frontiers | Research Topics

Population genetics and conservation of aquatic species

Edited by
Shaokui Yi, Yanhe Li, Cong Zeng and Narongrit Muangmai

Published in
Frontiers in Genetics
Frontiers in Ecology and Evolution

ISSN 1664-8714
ISBN 978-2-83251-335-4
DOI 10.3389/978-2-83251-335-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Population genetics and conservation of aquatic species

Topic editors
Shaokui Yi — Huzhou University, China
Yanhe Li — Huazhong Agricultural University, China
Cong Zeng — Shanghai Jiao Tong University, China
Narongrit Muangmai — Kasetsart University, Thailand

Citation
Yi, S., Li, Y., Zeng, C., Muangmai, N., eds. (2023). *Population genetics and conservation of aquatic species*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-335-4

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1052740

OPEN ACCESS

EDITED AND REVIEWED BY
Samuel A. Cushman,
Forest Service (USDA), United States

*CORRESPONDENCE
Cong Zeng,
congzeng@sjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 24 September 2022
ACCEPTED 16 December 2022
PUBLISHED 04 January 2023

CITATION
Yi S, Zeng C, Li Y and Muangmai N (2023), Editorial: Population genetics and conservation of aquatic species. *Front. Genet.* 13:1052740. doi: 10.3389/fgene.2022.1052740

Editorial: Population genetics and conservation of aquatic species

Shaokui Yi[1], Cong Zeng[2]*, Yanhe Li[3] and Narongrit Muangmai[4]

[1]School of Life Sciences, Huzhou University, Huzhou, China, [2]School of Oceanography, Shanghai Jiaotong University, Shanghai, China, [3]College of Fisheries, Huazhong Agricultural University, Wuhan, China, [4]Faculty of Fisheries, Kasetsart University, Bangkok, Thailand

KEYWORDS

aquatic species, population, genetics, conservation, structure

Editorial on the Research Topic
Population genetics and conservation of aquatic species

In recent decades, efforts to protect many terrestrial taxa have slowed their rates of extinction. Unfortunately, the outlook for a number of aquatic organisms, such as amphibians, corals, fish and other aquatic species is still not optimistic. Many aquatic species are highly threatened by anthropogenic and environmental disturbances, such as climatic change, overfishing, habitat elimination and fragmentation, and invasive species (Buchanan et al., 2016; Sowińska-Świerkosz and Kolejko, 2019). However, the conservation of these species requires knowledge of their spatial diversity and population structure, and the inaccessibility of aquatic animals poses a great challenge to traditional surveys. Population genetics provides the tools to describe genetic diversity within and among populations, while it also provides the basic theory for understanding the evolutionary change and resulting patterns of genetic variation in different populations. This information greatly contributes to the integrated concept of biodiversity conservation, which is needed to define the goals and methods of conservation programs and to set priorities (Loeschcke et al., 2013). Understanding the genetic landscape of natural populations is one of the key concerns for the development of conservation management strategies. Meanwhile, population genetic data for economically important aquatic species is also essential for the optimal utilization of this genetic resource in breeding programs. Importantly, with the great advances of sequencing technologies, detecting genomic variation (e.g., microsatellites, mitochondrial genes and single nucleotide polymorphisms) is becoming increasingly inexpensive and efficient. Molecular markers have been extensively applied in population genetics studies of aquatic animals during the past decade (Chapman et al., 2012; Yi et al., 2019). A large number of molecular markers can provide an efficient means to infer the population history and status of examined species and to predict future changes. The number and type of markers used are critical factors when planning a population genetic study. In some cases, the results generated with different traditional markers, such as mitochondrial genes and nuclear microsatellite markers (simple sequence repeats, SSRs), have been inconsistent (Baisvar et al., 2018; Wang et al., 2019; Zhong et al., 2019). With the aid of cost-effective genotyping technology, genome-wide single nucleotide polymorphism (SNP) markers could help us to obtain more reliable population genetic data, which is of great importance to complement or replace existing conservation strategies. For this Research Topic, we gathered studies of aquatic populations that use these powerful molecular markers to interpret population structure, phylogeography, or evolutionary processes. Their findings can be directly applied to conservation efforts. The outcomes of studies could directly provide suggestions or implements for conservation.

Thanks to the combined effort of all the Editors, we are pleased to present 11 papers authored by 80 excellent researchers from various fields. The paper by Huo et al. focuses on the genetic diversity and population structure of *Triplophysa tenuis*, an important indigenous fish in the Xinjiang Tarim River that is facing overfishing and habitat degradation. For the study, a large number of SNPs were obtained with the genotyping-by-sequencing (GBS) method, and the eight populations were found to have high levels of genetic diversity, with substantial genetic differentiation among populations. Hu et al. investigated the genetic differentiation of populations of a commercially important sleeper fish, *Odontobutis potamophilus*, using SNP markers, and suggest several conservation strategies in their report. The work by Repullés et al. evaluates the genetic structure and connectivity pattern of the endangered coral *Cladocora caespitosa* across its entire distribution range in the Mediterranean Sea. Their paper provides a better understanding of this endangered scleractinian coral, which allows for more informed conservation decisions. Most interestingly, Gilles et al. report massive introgressive hybridization between two distinct genera in Cyprinidae, and they suggest that the hybridization could generate a wide spectrum of hybrids that are a potential source of important evolutionary novelties. The paper by Collins et al. evaluates life history variation in the anadromous migration of *Oncorhynchus mykiss* using whole-genome resequencing, and report that a region on chromosome Omy12 may represent a minor effect gene. Also in this Research Topic, papers related to population genetics of several fish species are presented, including *Triplophysa robusta* (Zhong et al.), *Hippocampus erectus* (Luo et al.), *Gymnocypris przewalskii* (Fang et al.), *Coilia brachygnathus* (Zhai et al.), *Schizothorax biddulphi* (Nie et al.) and *Hemiculter leucisculus* (Gu et al.). In these studies, SNPs, SSRs or mitochondrial genes were applied to successfully identify patterns of genetic diversity, population structure and phylogeography. Their results should be very useful for assessing the population dynamics of these species and for developing future conservation strategies.

In conclusion, all of the papers in this Research Topic evaluate the population genetics of important aquatic species using molecular markers. The efforts of these researchers further the understanding of aquatic genetic resources and can help guide conservation and management strategies of aquatic species.

Author contributions

SY and CZ wrote the manuscript, YL and NM revised the manuscript.

Funding

This study was supported by the Public Welfare Applied Research Project of Huzhou Science and Technology Bureau (No. 2021GZ36).

Acknowledgments

We greatly appreciate the kind help of the journal's staffs and all reviewers in preparing this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Baisvar, V. S., Kumar, R., Singh, M., Singh, A. K., Chauhan, U. K., Mishra, A. K., et al. (2018). Genetic diversity analyses for population structuring in *Channa striata* using mitochondrial and microsatellite DNA regions with implication to their conservation in Indian waters. *Meta Gene* 16, 28–38. doi:10.1016/j.mgene.2018.01.004

Buchanan, J. R., Krupp, F., Burt, J. A., Feary, D. A., Ralph, G. M., and Carpenter, K. E. (2016). Living on the edge: Vulnerability of coral-dependent fishes in the Gulf. *Mar. Pollut. Bull.* 105 (2), 480–488. doi:10.1016/j.marpolbul.2015.11.033

Chapman, B. B., Skov, C., Hulthén, K., Brodersen, J., Nilsson, P. A., Hansson, L. A., et al. (2012). Partial migration in fishes: Definitions, methodologies and taxonomic distribution. *J. fish Biol.* 81 (2), 479–499. doi:10.1111/j.1095-8649.2012.03349.x

V. Loeschcke, J. Tomiuk, and S. K. Jain (Editors) (2013). *Conservation genetics* (Basel: Birkhäuser), 68, 3.

Sowińska-Świerkosz, B., and Kolejko, M. (2019). Extinction risk to lake minnow (Eupallasella percnurus) due to habitat loss: Eastern Poland case study. *Environ. Monit. Assess.* 191 (9), 571. doi:10.1007/s10661-019-7731-6

Wang, H., Wang, T., Li, W., and Liu, H. (2019). The genetic diversity, individual relatedness and possible mating system of an isolated population of the Cyprinid species Megalobrama pellgrini in upper reaches of the Changjiang (Yangtze) River, China. *J. Oceanol. Limnol.* 37 (3), 1042–1050. doi:10.1007/s00343-019-8152-7

Yi, S., Wang, W., and Zhou, X. (2019). Genomic evidence for the population genetic differentiation of *Misgurnus anguillicaudatus* in the Yangtze River basin of China. *Genomics* 111 (3), 367–374. doi:10.1016/j.ygeno.2018.02.011

Zhong, J., Yi, S., Ma, L., and Wang, W. (2019). Evolution and phylogeography analysis of diploid and polyploid *Misgurnus anguillicaudatus* populations across China. *Proc. Biol. Sci.* 286 (1901), 20190076. doi:10.1098/rspb.2019.0076

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 20 January 2022
doi: 10.3389/fgene.2021.830626

Potential Risks for Seahorse Stock Enhancement: Insight From the Declivity of Genetic Levels With Hatchery Management

*Wei Luo[1†], Qing Wu[1†], Xiaoyang Zhang[1], Yuling Wei[1], Min Liao[1], Tong Gao[1], Yibo Zhang[1], Shoudong Zhang[1], Pengyu Chen[1], Zhonggang Guo[2], Yinlin Xiong[3], Zhou Xu[4] and Zongjun Du[1]**

[1]College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China, [2]Chongzhou Agricultural and Rural Bureau, Chengdu, China, [3]The Original Stock Farm of Leiocassis Longirostris of Sichuan Province, Chengdu, China, [4]Mianyang Academy of Agricultural Sciences, Mianyang, China

Stock enhancement is one of the potential management strategies for the fishery. To better understand the impaction of stock enhancement, we simulated an experiment for lined seahorse (*Hippocampus erectus*) and evaluated the genetic structure after stock enhancement. In this study, we found the numbers of alleles (N_A) and heterozygosity (H_O) of stock enhancement strains were lower than those of the wild collections, while the inbreeding coefficient (F_{IS}) and relatedness index were higher. Within the 3 generations of stock enhancement strain, the N_A, H_O and polymorphism information content (*PIC*) didn't change significantly. In addition, the F_{ST} value indicated that the genetic differentiation between the stock enhancement strains and the first wild collection reached an intermediate level, which could lead to substructuring in wild populations. Overall, these findings revealed a potential genetic risk associated with the release of hatchery strains into wild populations.

Keywords: stock enhancement, genetic diversity, population genetics, genetic risk, seahorse

OPEN ACCESS

Edited by:
Shaokui Yi,
Huzhou University, China

Reviewed by:
Jing Chen,
Huazhong Agricultural University, China
Qian Gu,
Hunan Normal University, China

****Correspondence:***
Zongjun Du
14364@sicau.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *07 December 2021*
Accepted: *27 December 2021*
Published: *20 January 2022*

Citation:
Luo W, Wu Q, Zhang X, Wei Y, Liao M, Gao T, Zhang Y, Zhang S, Chen P, Guo Z, Xiong Y, Xu Z and Du Z (2022) Potential Risks for Seahorse Stock Enhancement: Insight From the Declivity of Genetic Levels With Hatchery Management. Front. Genet. 12:830626. doi: 10.3389/fgene.2021.830626

INTRODUCTION

Recent studies show that the living conditions of most wild commercial and protected fisheries have been altered by overfishing, pollution and degradation of habitats, and leading to the decline of the stocks (Hansen, 2002). Supporting weak wild populations through the release of conspecifics (stock enhancement) is being used increasingly in conservation practice. More than 65 marine and brackish-water species have been extensively subjected to stocking practices across 27 countries with the goal of increasing exploitable resources (Bartley and Bell, 2008).

Stock enhancement has been used for more than 100 years; however, its efficacy and risk to natural populations has been controversial (Stottrup and Sparrevohn, 2007; Araki and Schmid, 2010). In particular, hatchery often leads to genetic, morphological and behavioral differences between captive-bred and wild populations due to environmental differences, and human intervention (Svasand et al., 2000; Araki et al., 2007; Blanchet et al., 2008). Genetic effects of hatchery programs on wild populations have been investigated in many species, but the effects affect the ecology, genetics, and fitness of the wild stocks is still not clear (Araki and Schmid, 2010; Hold et al., 2013). Some researchers did not find any direct evidence for the genetic effects of hatchery fish on wild populations, such as in sea trout (*Salmo trutta*) (Palm et al., 2003), *G. morhua* (Svasand et al.,

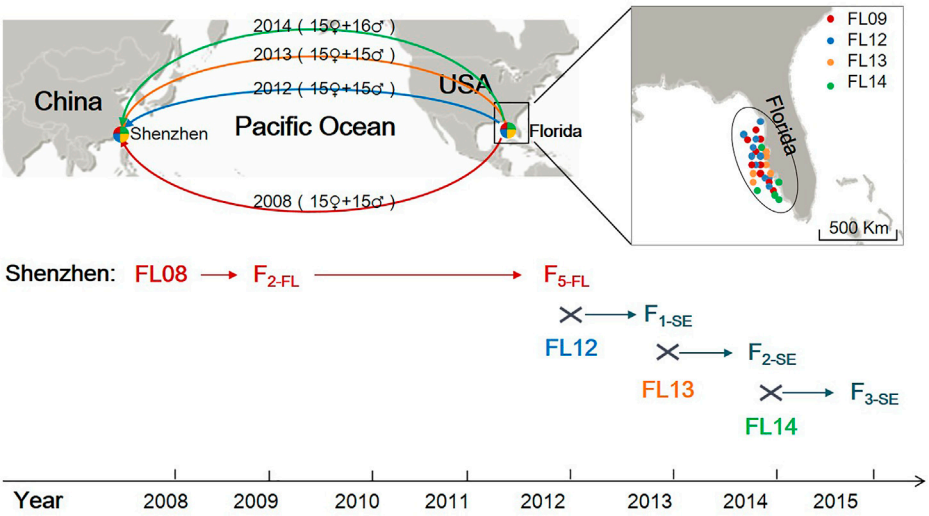


FIGURE 1 | The sampling sites of wild lined seahorses, a sketch of the introduction from the United States to China and a flowchart of the hatchery and a simulated stock enhancement in China. Red, blue, orange, and green lines indicate the introduction from Florida to Shenzhen in 2008, 2012, 2013, and 2014, respectively.

2000), channel catfish (*Ictalurus punctatus*) (Simmons et al., 2006). Conversely, other researchers reported that hatchery fish have obvious genetic effects on wild populations, such as in Atlantic salmon (Skaala et al., 2006), coho salmon (Eldridge and Naish, 2007). In such cases, evaluating the utility of captive breeding for stock enhancement considering the associated genetic risks is a prerequisite for future management decisions.

Seahorses, genus *Hippocampus*, are highly unusual marine fishes with unique body morphology, and specialized life history traits. Their biological and ecological characteristics, including low mobility, small home ranges, sparse distribution, male pregnancy, lengthy parental care and strict monogamy in most species, might render them vulnerable to overfishing or habitat damage (Foster and Vincent, 2004; Curtis and Vincent, 2006). Worldwide, many seahorse species are considered to be under threat of overexploitation for traditional medicines, aquarium trade and ornaments (Foster and Vincent, 2004; Koldewey and Martin-Smith, 2010). The entire genus *Hippocampus* is listed on Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), and the majority of seahorse species are listed as data deficient with several others listed as vulnerable or endangered (Pollom, 2017).

In this study, we simulated a stock enhancement and evaluated the impact of hatchery rearing on the genetic structure of wild seahorse populations using eleven microsatellite markers, with the purpose of providing a scientific basis for wild resource recovery for seahorse in the future.

MATERIALS AND METHODS

Study Populations

Four collections of the lined seahorses (FL08, FL12, FL13, and FL14) were sampled from the region of the west coast of Florida (United States) in 2008, 2012, 2013, and 2014, respectively

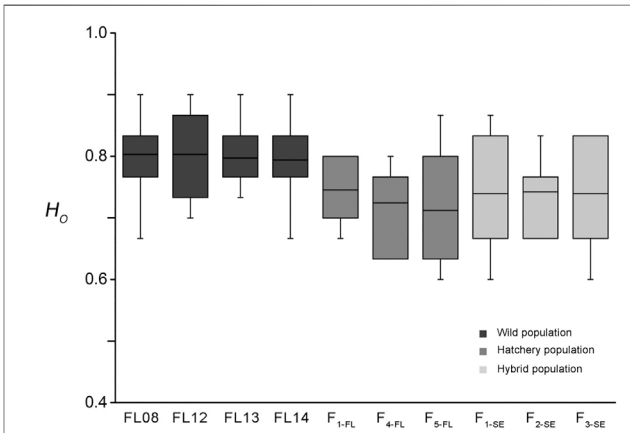


FIGURE 2 | The box plot of H_O values of wild (FL08, FL12, FL13, and FL14), hatchery ($F_{1\text{-}FL}$, $F_{4\text{-}FL}$, and $F_{5\text{-}FL}$) and "enhancement stock" strains ($F_{1\text{-}SE}$, $F_{2\text{-}SE}$, and $F_{3\text{-}SE}$) of lined seahorse.

(**Figure 1**). The seahorses were cultured in separate concrete outdoor ponds (3 × 3 × 1 m), with recirculating sea water treated with double sand filtration. Seahorses were fed daily with rotifers, copepods, *artemia*, *Mysis* spp., and *Acetes* spp. The FL08 collection, as initial breeders, was used for artificial breeding. Their fry ($F_{1\text{-}FL}$) were immediately transferred to another tank after they were hatched out. After about 6 months of feeding, they reached maturity and 30 individuals (sex ration 1:1) were randomly selected for reproducing the next generation. In a similar fashion, offspring of the second to the fifth generation of FL08 were obtained. In 2012, to simulate the influence of stock enhancement in releasing the hatchery strains into the wild, we designed a mating between the FL12 collection and the fifth generation of FL08 ($F_{5\text{-}FL}$). We randomly selected 30 individuals of $F_{5\text{-}FL}$ (sex ration 1:1) and put them together with FL12

collection (sex ration 1:1) in a culture pond. In 2013 and 2014, we did a mating between $F_{1\text{-SE}}$ (randomly selected) and FL13 and $F_{2\text{-SE}}$ (randomly selected) and FL14, respectively. In this study, we assumed that the "stock enhancement" strains (named $F_{i\text{-SE}}$, where *i* is the generation) were mated with wild individuals with number ratio of 1:1 each generation in the simulation. In this study, thirty individuals of each collection, including 4 wild collections, 2 hatchery strains, and 3 "stock enhancement" strains were sampled for genetic analysis.

DNA Extraction, Amplification and Microsatellite Analysis

Genomic DNA was extracted from approximately 100 mg of dorsal fin tissue from every sample by standard phenol-chloroform protocols. All the samples were genotyped with eleven microsatellite markers, of which *Hier-ssr3*, *Hier-ssr7*, *Hier-ssr8*, *Hier-ssr9*, *Hier-ssr10*, *Hier-ssr13*, *Hier-ssr15*, *Hier-ssr17*, *Hier-ssr28,* and *Hier-ssr29* were previously developed by our group (Luo et al., 2015; Arias et al., 2016), and *Hier-ssr51* was designed in this study (**Supplementary Table S1**). The method of PCR amplification and microsatellite analysis referred to Luo et al. (2017).

Genetic Variation Analysis

The number of alleles (N_A), the effective number of alleles (N_{AE}), allele frequency (*AF*), observed heterozygosity (H_O), expected heterozygosity (H_E), the genetic distance (GD), and the departure from Hardy-Weinberg equilibrium (*HWE*) of each locus were calculated using POPGENE32 (ver. 1.32). A UPGMA phylogenetic tree was built using MEGA 5.1 software according to the information on *GD* between different populations. Analysis of molecular variance (AMOVA) of the genetic structure, the pairwise coefficient of genetic differentiation (F_{ST}), the coefficient of inbreeding (F_{IS}) among the ten populations, and the presence of null alleles were estimated using Arlequin software (ver. 3.11). The relatedness index among the individuals sampled was calculated using the method of TrioML by Coancestry software.

RESULTS

Genetic Diversity of Wild, Hatchery and "Stock Enhancement" Strains

A total of 103 alleles at all loci were detected in the 4 wild collections with a mean value of 9.36 per locus, and there was no obvious change of N_A within the 4 collections. The mean N_A value of first hatchery strain generation of FL08 ($F_{1\text{-FL}}$) was 8.18 per locus, and that of the fifth generation ($F_{5\text{-FL}}$) declined to 8.00. The N_A of the first generation of stock enhancement strains ($F_{1\text{-SE}}$) recovered to 8.55, and that of the second ($F_{2\text{-SE}}$) and the third generation ($F_{3\text{-SE}}$) of stock enhancement strains was a little lower. Similar to N_A, the N_{AE} values of the hatchery strains gradually declined over generations, and that of the 3 "stock enhancement" strains slightly recovered, but it was still lower than that of wild collections (**Supplementary Table S2**).

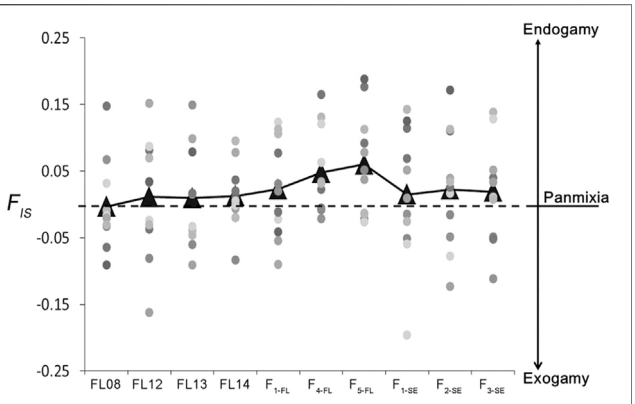


FIGURE 3 | The inbreeding coefficient (F_{IS}) of lined seahorse collections (4 wild collections, 3 hatchery strains and 3 "stock enhancement" strains). Dots with different colors represent different microsatellite loci. Triangles represent the mean values of F_{IS} of different collections.

No null alleles were detected in the hatchery strains and stock enhancement strains. Linkage disequilibrium was assessed and no significant departure from equilibrium levels was detected in any sample. No locus was significantly departed from *HWE* ($p < 0.05$) in the 4 wild collections, while 1, 2, and 2 were significantly departed from *HWE* ($p < 0.05$) in $F_{1\text{-FL}}$, $F_{2\text{-FL}}$, and $F_{5\text{-FL}}$ and 1, 1 and 2 were in $F_{1\text{-SE}}$, $F_{2\text{-SE}}$, and $F_{3\text{-SE}}$, respectively. All significant deviations from Hardy–Weinberg expected heterozygosities were due to an excess of homozygotes. Compared with the genotype data of the FL08 and $F_{5\text{-FL}}$ collections, some alleles, especially rare alleles of *Hier-ssr7*, *Hier-ssr8*, *Hier-ssr9,* and *Hier-ssr10*, were found to be lost in the process of captive breeding. Similar to the hatchery strains, some alleles of *Hier-ssr9* and *Hier-ssr10* were also lost in the "stock enhancement" strains, but some alleles, such as those of *Hier-ssr15*, and reappeared in these strains (**Supplementary Table S2**).

The minimal allele frequency (*MAF*) exhibited a growing disparity with generations due to allele loss; some alleles were gradually reduced until they disappeared, and some gradually increased and dominated the population. The mean *MAF* values of the 4 wild collections over all loci were 0.023, 0.026, 0.029, and 0.021; while those of the hatchery strains and "stock enhancement" strains increased, especially for $F_{5\text{-FL}}$, with mean values of 0.036.

The mean value of H_O of the FL08 collection over all loci was 0.80, with a minimum of 0.67 at locus *Hier-ssr38* and a maximum of 0.90 at locus *Hier-ssr26* and *Hier-ssr15*. The mean values of H_O of other 3 wild collections were similar with that of FL08. While that of $F_{1\text{-FL}}$, $F_{4\text{-FL}}$, and $F_{5\text{-FL}}$ was 0.75, 0.72, and 0.71, respectively. The mean values of H_O of the 3 "stock enhancement" strains were all about 0.74, which were higher than that of $F_{5\text{-FL}}$ but lower than those of wild collections (**Figure 2**). The value of H_E of wild collections ranged from 0.80 to 0.81, while that in hatchery strains from 0.75 to 0.76 and "stock enhancement" strains from 0.74 to 0.76 (**Supplementary Table S2**).

The mean *PIC* values of all loci in the 4 wild collections were closed to 0.76. While, the mean *PIC* values of $F_{1\text{-FL}}$, $F_{4\text{-FL}}$, and $F_{5\text{-}}$

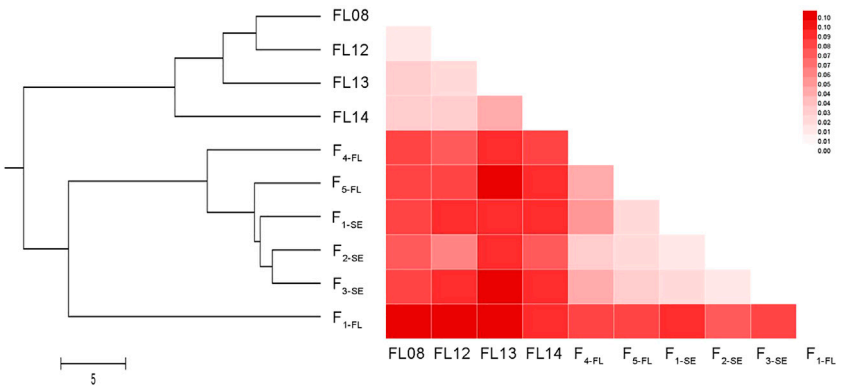


FIGURE 4 | The UPGMA phylogenetic tree and heatmap of F_{ST} values among 10 lined seahorse collections. UPGMA clustering was based on *GD* among pairs of populations and the tree was built using MEGA 5.1 software.

$_{FL}$ strains was 0.71, 0.69, and 0.69, respectively. The *PIC* values of "stock enhancement" strains ranged from 0.71 to 0.74, which were higher than those of hatchery strains but lower than those of wild collections (**Supplementary Table S2**).

The inbreeding coefficient (F_{IS}) was close to 0 for the 4 wild collections. The F_{IS} values of the 3 hatchery strains were higher than those of wild collections, and the inbreeding rate gradually accumulated over generations with highest F_{IS} value of 0.06 in the $F_{5\text{-FL}}$ strain. In addition, the F_{IS} values of all the 3 "stock enhancement" strains maintained about 0.02 (**Figure 3**), indicating there existed a small risk of inbreeding in the wild after stock enhancement. The mean relatedness index of the wild collections was from 0.067 to 0.070; while that of the hatchery stains was higher and increased over generations (the relatedness index of $F_{5\text{-FL}}$ reached to 0.106). The mean relatedness index of "stock enhancement" strains was lower than that of hatchery stains but higher than that of wild collections (Appendix A).

Genetic Divergence Among Wild, Hatchery and "Stock Enhancement" Collections

The range of *GD* of the ten collections was from 0.08 (the pair of $F_{1\text{-SE}}$ and $F_{2\text{-SE}}$) to 0.57 (the pair of $F_{2\text{-SE}}$ and FL12). The UPGMA phylogenetic tree based on *GD* among pairs of populations (**Figure 4**) was divided into two branches: one represents the 4 wild collections, and the other represents the hatchery strains and "stock enhancement" strains. In the lower branch, the 3 "stock enhancement" strains were clustered together firstly, and then $F_{1\text{-FL}}$, $F_{4\text{-FL}}$, and $F_{5\text{-FL}}$ was separated from the branch of "stock enhancement" strains successively.

The value of F_{ST} ranged from 0.02 (the pair of FL08 and FL13) to 0.11 (the pair of FL12 and $F_{1\text{-SE}}$) for all collections (**Figure 4**). The F_{ST} values between $F_{3\text{-SE}}$ and FL08, FL12, FL13, and FL14 were 0.11, 0.11, 0.10, and 0.09, respectively, indicating a medium level of genetic differentiation between "stock enhancement" strains and wild populations. The result of AMOVA of all ten collections revealed that 92.65% of genetic differentiation was within populations, and only 7.35% was among collections (**Supplementary Table S3**).

DISCUSSION

In aquaculture and wildlife protection, understanding whether genetic variability is lost in cultured stocks after long periods of hatchery production is important, because reduction in variability may possibly result in the loss of genetic variation for disease resistance and reduce a population's capability to adapt to new environments (Li et al., 2004). Unfortunately, reduction in allelic richness, which was found in many reported cultured species, including fish, shrimp and shellfish (Li et al., 2004; Kohlmann et al., 2005; Marchant et al., 2009), and seems to be a characteristic of hatchery populations. There is no exception in the captive breeding of seahorse. The results showed that the values of N_A, H_O, and *PIC* of the 3 hatchery strains were lower than those of wild collections, and the impact of hatchery rearing on the loss of allele richness seemed to be more and more serious over generations. Specifically, the $F_{5\text{-FL}}$ strain lost approximately 13% of the number of alleles and 11% of the heterozygosity compared to FL08 collection, respectively. In addition, we found the mean relatedness index of hatchery strains was higher than that of wild collections, especially some pairs of hatchery strains had high relatedness value. Serial founder effects across generation and associated genetic drift could be the major factor responsible for the reduction of diversity of hatchery strains. Moreover, unequal contribution to offspring of breeders due to artificial selection in the hatchery process, may lead to a reduction of some alleles and even the disappearance and increase of the effect of genetic drift (Nakajima et al., 2014).

The reduction in genetic diversity and changes in allelic frequencies of hatchery strains may pose potential genetic risks to wild populations in supportive breeding programs (Machado-Schiaffino et al., 2007). The introduction of progeny with low genetic diversity into wild seahorse populations has led to lowered overall genetic variability of the resulting populations (Utter, 1998). In our study, the values of N_A, H_O, and *PIC* of the 3 "stock enhancement" strains were lower, but *MAF* were higher than those of wild collections. In addition, the allelic richness and heterozygosity of "stock enhancement" strains did not rise within

3 generations after seahorse being released into wild, indicating consecutive interbreeding with wild-collected seahorses failed to reintroduce lost alleles or genetic diversity. The reduction of genetic diversity in wild population seems to persist for a long time. Specially, seahorse has high mate fidelity and lengthy parental care (Foster and Vincent, 2004; Koldewey and Martin-Smith, 2010), restricting multiple mating between individuals during their lifespan; they also have low mobility and small home ranges (Curtis and Vincent, 2006; Koldewey and Martin-Smith, 2010), limiting gene flow amongst wild populations. Thus, their specialized life history traits may make seahorse populations particularly susceptible to anthropogenic disturbances, and their genetic diversity is easy to reduce and difficult to recover (Foster and Vincent, 2004; Curtis and Vincent, 2006). Our findings are in accordance with numerous previous reports on stock enhancement of fish species, such as brown trout (*Salmo trutta*) (Hansen, 2002), red sea bream (*Pagrus major*) (Kitada et al., 2009), and Pacific herring (*Clupea pallasii*) (Sugaya et al., 2008).

In this study, some alleles were lost and the allele frequency changed a lot in the hatchery process, which lead to an increase in the mean value of *MAF*, and a significant departure from *HWE* for some loci in the hatchery and "stock enhancement" strains. A previous study reported on pacific abalone (*Haliotis discus hannai*) that random changes in allele frequencies could be caused by genetic drift in a depleted or small hatchery population (Li et al., 2004). The non-random mating system has changed the overall allelic composition of the farmed strain relative to the wild population, and the overall allelic composition of the "stock enhancement" strains might not be possible to restore to the original form (Li et al., 2004; Kitada et al., 2009; Hold et al., 2013). The concern is that declines in variation at neutral markers such as these may be indicative of a loss of variation at coding regions, which is the source of variation in important commercial traits such as growth rate and disease resistance.

Selection that occurs in hatchery rearing could be deleterious because traits that are advantageous in the reared environment may be disadvantageous in the wild. In other words, if the trait distribution in a wild population is at an optimum that has been shaped by selection in the wild environment, releasing individuals into the population that have a different distribution as a result of selection in a hatchery will result in a reduction in the mean fitness of the population (Araki et al., 2007; Kitada et al., 2009; Jonsson et al., 2016). Factors such as predation, escape, competition and stress tolerance are less important under culturing environments with adequate food, less competition, a steady water environment, and a homogeneous habitat in the long-term process of hatching (Stottrup and Sparrevohn, 2007; Kitada et al., 2009). Experience with salmonid culture shows that, with or without intentional selection of parents, each successive generation of captive brood stock will favor those parents and progeny that are most amenable to conditions of culture (Utter, 1998); while, in the wild environment, populations have evolved different life-history strategies, which can be attributed to interior intra-species interactions, and environment heterogenetity (Svasand et al., 2000). A potential risk associated with inbreeding depression could be brought into the natural ecosystem because of high inbreeding rate and relatedness value accumulated over generations in hatchery rearing. Even after serial generations of mating with wild individuals, our result demonstrated the "stock enhancement" strains still kept medium inbreeding rate and relatedness. The phenomenon was also found in most cultured species, such as abalone (*Haliotis discus hannai*) (Marchant et al., 2009) and common carp (*Cyprinus carpio* L.) (Kohlmann et al., 2005). The tendency towards homozygosity could be explained by the founder effect (Marchant et al., 2009).

Recent studies demonstrated that repeated interbreeding between a natural population and mass-released hatchery strains could result in a non-random mating system or, possibly, assortative mating that could generate within-population substructuring (Wahlund effect), and inbreeding (Sekino et al., 2005). In this study, the pairwise F_{ST} and *GD* values indicated a medium genetic differentiation between "stock enhancement" strains and wild collections (According to Wright's criteria ($0 < F_{ST} < 0.05$, no differentiation; $0.05 < F_{ST} < 0.15$, medium level; and $0.15 < F_{ST} < 0.25$, high level). The results of AMOVA supported relatively high genetic variation within collections. The genetic differentiation between "stock enhancement" strains and wild collections was medium, which poses a genetic risk of within-population substructuring to the natural populations (Sekino et al., 2005).

Although a potential risk exists, many measures can be used to mitigate or remove the risk for seahorse stock enhancement. Introducing large numbers of wild breeders into hatchery could be one of the most effective ways to maintain the genetic diversity of hatchery populations (Kitada et al., 2009). Moreover, it is necessary for us to maintain clear genetic lineage by building marker-based pedigree assignment and to design a mating strategy based on genetic distance to avoid inbreeding depression in seahorse hatcheries.

In conclusion, the results of this study suggested that the genetic diversity in terms of allelic diversity and heterozygosity decreased but the inbreeding coefficient increased over generations in the hatchery compared to wild collections. Moreover, the genetic diversity of "stock enhancement" collections was lower than the wild population and it seemed that consecutive interbreeding with wild-collected seahorses did not improve the genetic diversity very well. In addition, the genetic differentiation between wild and "stock enhancement" collections reached a medium level, and which could lead to substructuring in wild populations. Considering the highly unique life history traits of the seahorse, these findings reveal a potential risk associated with the release of hatchery fishes into wild populations. This study provides valuable information on seahorse resource conservation and recovery from a genetic perspective.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by Sichuan Agricultural University.

AUTHOR CONTRIBUTIONS

WL, JZ, and QW conceived and designed the experiments. QW and ZX analyzed the data. ZX, YW, ML, TG, YZ, SZ, PC, ZG, YX, and XZ contributed reagents/materials/analysis tools. WL and QW wrote the paper. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (41706171).

ACKNOWLEDGMENTS

We greatly appreciate Prof. Junda Lin of Florida Institute of Technology and Prof. Qiang Lin of South China Sea Institute of Oceanology, CAS for collecting the seahorse samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.830626/full#supplementary-material

REFERENCES

Araki, H., Cooper, B., and Blouin, M. S. (2007). Genetic Effects of Captive Breeding Cause a Rapid, Cumulative Fitness Decline in the Wild. *Science* 318, 100–103. doi:10.1126/science.1145621

Araki, H., and Schmid, C. (2010). Is Hatchery Stocking a Help or Harm? *Aquaculture* 308, S2–S11. doi:10.1016/j.aquaculture.2010.05.036

Arias, M. C., Aulagnier, S., Baerwald, E. F., Barclay, R. M., Batista, J. S., Beasley, R. R., et al. (2016). Microsatellite Records for Volume 8, Issue 1. *Conserv. Genet. Resour.* 8, 43–81. doi:10.1007/s12686-016-0522-2

Bartley, D. M., and Bell, J. D. (2008). Restocking, Stock Enhancement, and Sea Ranching: Arenas of Progress. *Rev. Fish. Sci.* 16, 357–365. doi:10.1080/10641260701678058

Blanchet, S., Páez, D. J., Bernatchez, L., and Dodson, J. J. (2008). An Integrated Comparison of Captive-Bred and Wild Atlantic salmon (*Salmo salar*): Implications for Supportive Breeding Programs. *Biol. Conservation* 141, 1989–1999. doi:10.1016/j.biocon.2008.05.014

Curtis, J. M. R., and Vincent, A. C. J. (2006). Life History of an Unusual marine Fish: Survival, Growth and Movement Patterns of *Hippocampus guttulatus* Cuvier 1829. *J. Fish. Biol.* 68, 707–733. doi:10.1111/j.0022-1112.2006.00952.x

Eldridge, W. H., and Naish, K. A. (2007). Long-term Effects of Translocation and Release Numbers on fine-scale Population Structure Among Coho salmon (*Oncorhynchus kisutch*). *Mol. Ecol.* 16, 2407–2421. doi:10.1111/j.1365-294x.2007.03271.x

Foster, S. J., and Vincent, A. C. J. (2004). Life History and Ecology of Seahorses: Implications for Conservation and Management. *J. Fish Biol.* 65, 1–61. doi:10.1111/j.0022-1112.2004.00429.x

Hansen, M. M. (2002). Estimating the Long-Term Effects of Stocking Domesticated trout into Wild Brown trout (*Salmo trutta*) Populations: an Approach Using Microsatellite DNA Analysis of Historical and Contemporary Samples. *Mol. Ecol.* 11, 1003–1015. doi:10.1046/j.1365-294x.2002.01495.x

Hold, N., Murray, L. G., Kaiser, M. J., Hinz, H., Beaumont, A. R., and Taylor, M. I. (2013). Potential Effects of Stock Enhancement with Hatchery-Reared Seed on Genetic Diversity and Effective Population Size. *Can. J. Fish. Aquat. Sci.* 70, 330–338. doi:10.1139/cjfas-2012-0378

Pollom, R. (2017). Hippocampus erectus. *The IUCN Red List of Threatened Species 2017: e.T10066A20191442*. doi:10.2305/IUCN.UK.2017-3.RLTS.T10066A20191442.en

Jonsson, P. R., Corell, H., André, C., Svedäng, H., and Moksnes, P.-O. (2016). Recent Decline in Cod Stocks in the North Sea-Skagerrak-Kattegat Shifts the Sources of Larval Supply. *Fish. Oceanogr.* 25, 210–228. doi:10.1111/fog.12146

Kitada, S., Shishidou, H., Sugaya, T., Kitakado, T., Hamasaki, K., and Kishino, H. (2009). Genetic Effects of Long-Term Stock Enhancement Programs. *Aquaculture* 290, 69–79. doi:10.1016/j.aquaculture.2009.02.011

Kohlmann, K., Kersten, P., and Flajšhans, M. (2005). Microsatellite-based Genetic Variability and Differentiation of Domesticated, Wild and Feral Common Carp (*Cyprinus carpio* L.) Populations. *Aquaculture* 247, 253–266. doi:10.1016/j.aquaculture.2005.02.024

Koldewey, H. J., and Martin-Smith, K. M. (2010). A Global Review of Seahorse Aquaculture. *Aquaculture* 302, 131–152. doi:10.1016/j.aquaculture.2009.11.010

Li, Q., Park, C., Endo, T., and Kijima, A. (2004). Loss of Genetic Variation at Microsatellite Loci in Hatchery Strains of the Pacific Abalone (*Haliotis Discus Hannai*). *Aquaculture* 235, 207–222. doi:10.1016/j.aquaculture.2003.12.018

Luo, W., Qu, H., Li, J., Wang, X., and Lin, Q. (2015). A Novel Method for the Identification of Seahorses (Genus *Hippocampus*) Using Cross-Species Amplifiable Microsatellites. *Fish. Res.* 172, 318–324. doi:10.1016/j.fishres.2015.07.017

Luo, W., Wang, W.-M., Wan, S.-M., Lin, Q., and Gao, Z.-X. (2017). Assessment of Parental Contribution to Fast- and Slow-Growth Progenies in the blunt Snout Bream (Megalobrama amblycephala) Based on Parentage Assignment. *Aquaculture* 472, 23–29. doi:10.1016/j.aquaculture.2016.07.003

Machado-Schiaffino, G., Dopico, E., and Garcia-Vazquez, E. (2007). Genetic Variation Losses in Atlantic salmon Stocks Created for Supportive Breeding. *Aquaculture* 264, 59–65. doi:10.1016/j.aquaculture.2006.12.026

Marchant, S., Haye, P. A., Marín, S. A., and Winkler, F. M. (2009). Genetic Variability Revealed with Microsatellite Markers in an Introduced Population of the abaloneHaliotis Discus hannaiIno. *Aquac. Res.* 40, 298–304. doi:10.1111/j.1365-2109.2008.02089.x

Nakajima, K., Kitada, S., Habara, Y., Sano, S., Yokoyama, E., Sugaya, T., et al. (2014). Genetic Effects of marine Stock Enhancement: a Case Study Based on the Highly Piscivorous Japanese Spanish Mackerel. *Can. J. Fish. Aquat. Sci.* 71, 301–314. doi:10.1139/cjfas-2013-0418

Palm, S., Dannewitz, J., Järvi, T., Petersson, E., Prestegaard, T., and Ryman, N. (2003). Lack of Molecular Genetic Divergence between Sea-ranched and Wild Sea trout (Salmo trutta). *Mol. Ecol.* 12, 2057–2071. doi:10.1046/j.1365-294x.2003.01892.x

Sekino, M., Saido, T., Fujita, T., Kobayashi, T., and Takami, H. (2005). Microsatellite DNA Markers of Ezo Abalone (*Haliotis Discus Hannai*): a Preliminary Assessment of Natural Populations Sampled from Heavily Stocked Areas. *Aquaculture* 243, 33–47. doi:10.1016/j.aquaculture.2004.10.013

Simmons, M., Mickett, K., Kucuktas, H., Li, P., Dunham, R., and Liu, Z. (2006). Comparison of Domestic and Wild Channel Catfish (*Ictalurus punctatus*) Populations Provides No Evidence for Genetic Impact. *Aquaculture* 252, 133–146. doi:10.1016/j.aquaculture.2005.11.006

Skaala, Ø., Wennevik, V., and Glover, K. A. (2006). Evidence of Temporal Genetic Change in Wild Atlantic salmon, *Salmo salar* L., Populations Affected by Farm Escapees. *Ices J. Mar. Sci.* 63, 1224–1233. doi:10.1016/j.icesjms.2006.04.005

Stottrup, J. G., and Sparrevohn, C. R. (2007). Can Stock Enhancement Enhance Stocks? *J. Sea Res.* 57, 104–113. doi:10.1016/j.seares.2006.09.005

Sugaya, T., Sato, M., Yokoyama, E., Nemoto, Y., Fujita, T., Okouchi, H., et al. (2008). Population Genetic Structure and Variability of Pacific Herring *Clupea pallasii* in the Stocking Area along the Pacific Coast of Northern Japan. *Fish. Sci.* 74, 579–588. doi:10.1111/j.1444-2906.2008.01561.x

Svåsand, T., Kristiansen, T. S., Pedersen, T., Salvanes, A. G. V., Engelsen, R., Naevdal, G., et al. (2000). The Enhancement of Cod Stocks. *Fish Fish* 1, 173–205. doi:10.1046/j.1467-2979.2000.00017.x

Utter, F. (1998). Genetic Problems of Hatcher-Reared Progeny Released into the Wild, and How to deal with Them. *B. Mar. Sci.* 62, 623–640.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 03 March 2022
doi: 10.3389/fgene.2022.860678

Population Structure, Genetic Diversity and Differentiation of *Triplophysa tenuis* in Xinjiang Tarim River

Bin Huo[1], *Xuan Liu*[1], *Shengao Chen*[2], *Jieya Liu*[2], *Qiong Zhou*[1], *Jianzhong Shen*[1], *Dapeng Li*[1], *Rong Tang*[1], *Jing Chen*[1]* *and Xiaoyun Zhou*[1]*

[1]College of Fisheries, Huazhong Agricultural University, Wuhan, China, [2]College of Animal Science, Tarim University, Alar, China

Triplophysa tenuis is an important indigenous fish in the Xinjiang Tarim River. In this study, we collected 120 *T. tenuis* individuals from 8 *T. tenuis* populations in the Tarim River. Through genotyping-by-sequencing (GBS), a total of 582,678,756 clean reads were generated for all the genotypes, and after quality filtering, 595,379 SNPs were obtained for the population genetic analyses. Multiple genetic parameters showed that the 8 *T. tenuis* populations had high genetic diversity. Phylogenetic tree analysis indicated that all *T. tenuis* individuals were divided into five branches, the individuals from the north of Tarim River were grouped into cluster 1 (SF and WS) and cluster 3 (DWQ, TKX, and KZE), while the AETS, WLWT and LF individuals from the south of Tarim River were clustered into cluster 2. The result was consistent with the admixture analysis, which supported that the 8 *T. tenuis* populations were clustered into three subgroups. Furthermore, the pairwise F_{ST} values and genetic distance indicated that there was a large genetic differentiation between WS and other *T. tenuis* populations. Collectively, this study provides valuable genome-wide data for the conservation of natural *T. tenuis* populations in the Tarim River.

Keywords: Triplophysa tenuis, tarim river, population structure, genetic diversity, genetic differentiation

OPEN ACCESS

Edited by:
Cong Zeng,
Shanghai Jiao Tong University, China

Reviewed by:
Zongjun Du,
Sichuan Agricultural University, China
Baoying Guo,
Zhejiang Ocean University, China

***Correspondence:**
Jing Chen
cjing24511@163.com
Xiaoyun Zhou
zhouxy@mail.hzau.edu.cn

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *23 January 2022*
Accepted: *17 February 2022*
Published: *03 March 2022*

Citation:
Huo B, Liu X, Chen S, Liu J, Zhou Q, Shen J, Li D, Tang R, Chen J and Zhou X (2022) Population Structure, Genetic Diversity and Differentiation of Triplophysa tenuis in Xinjiang Tarim River. Front. Genet. 13:860678. doi: 10.3389/fgene.2022.860678

INTRODUCTION

Triplophysa tenuis, which belongs to *Triplophysa*, Nemacheilidae, and Cypriniformes, inhabits the backwater of rivers over sandy or muddy bottoms. It is recognized as an important indigenous fish with certain economic value in southern Xinjiang Tarim River, such as Kaidu River, Aksu River (Xie et al., 2015). The feeding habit of *T. tenuis* is omnivorous, partial to carnivorous, mainly feeding on benthos, fish, shrimp and insect larvae, followed by algae and organic debris (Tursun et al., 2005). The genus *Triplophysa* (Plateau loach) is a special group in the Qinghai-Tibet Plateau, which has strong adaptability to the highland environment. In recent years, affected by natural and anthropic factors, the Tarim River has experienced runoff curtailment and river desiccation, accompanied by river salinization, biodiversity decrease, and ecosystem service loss (Chen et al., 2011). As a result, the population size of *T. tenuis* has been diminishing and it is critically endangered due to its habitat limited to the upstream of the Tarim River. Although previous studies have reported the reproductive, morphological, and other biological characteristics of *T. tenuis*, the genetic diversity and genetic differentiation of *T. tenuis* populations are still unclear (Tursun et al., 2005; Liu et al., 2021). Understanding the genetic diversity and genetic differentiation among *T. tenuis* populations is essential for designing conservation and management strategies.

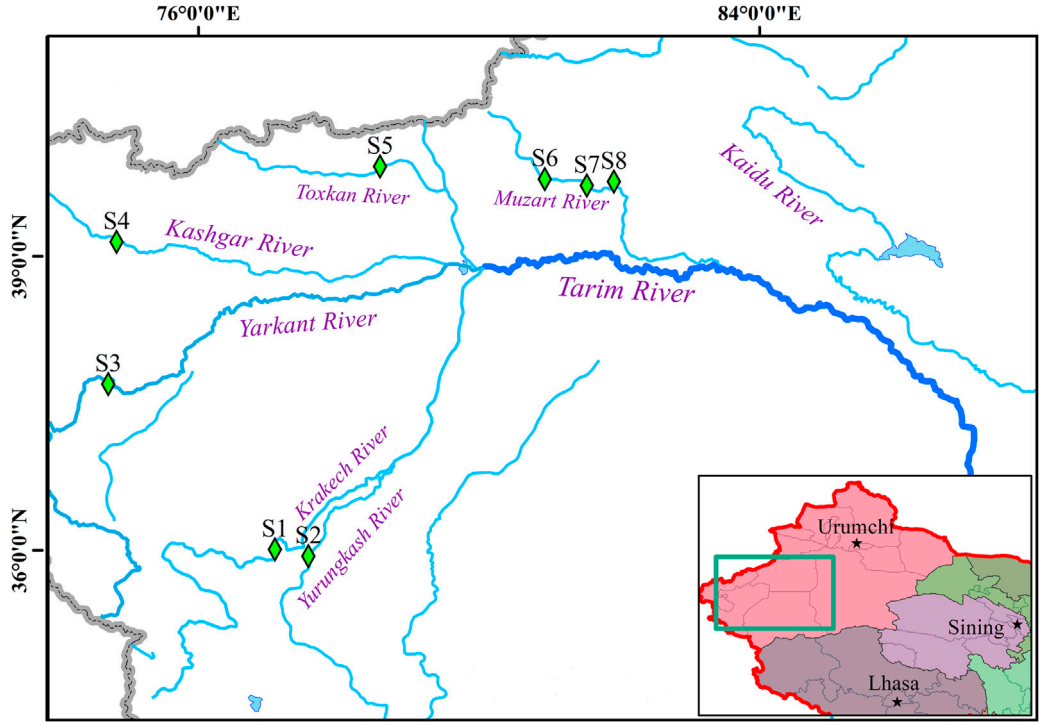


FIGURE 1 | The distribution of *T. tenuis* populations in the branches of the Tarim River. The blue lines represent the rivers. The green square symbols indicate the sampling locations. S1, WLWF; S2, LF; S3, AETS; S4, SF; S5, WS; S6, DWQ; S7, TKX; S8, KZE.

Traditional molecular markers play a crucial role in population genetics analysis, improving the understanding of complex quantitative traits, and facilitating marker-assisted breeding (Zhong et al., 2019). In the past decades, the most commonly used molecular markers include random amplified polymorphic DNA (RAPD), mitochondrial DNA sequences, simple sequence repeat (SSR), and single nucleotide polymorphism (SNP) (Younis et al., 2020). Notably, SNPs have become the preferred markers for genetic studies due to their unique characteristics, such as unbiased distribution, biallelic properties, and availability in the whole genome (Islam et al., 2021). Next-generation sequencing (NGS) technologies have been recently conducted for molecular marker development, population genetic analysis, and molecular breeding. For the species without a reference genome, reduced-representation genome sequencing (RRGS) can be utilized to obtain genome-wide genetic variation information. According to the different library construction strategies, RRGS is divided into reduced-representation libraries (RRL), restriction-site associated DNA (RAD), and genotyping-by-sequencing (GBS). Among them, GBS method reduces genome complexity and allows the discovery of genome-wide SNPs with a lower error rate based on restriction enzymes (REs) (Wang et al., 2020). GBS is the most widely used in aquatic species based on NGS with high-throughput genotyping and low per-sample, such as *Oryzias latipes* (Katsumura et al., 2019), *Cynoglossus semilaevis* (Zhang et al., 2020), and *Misgurnus anguillicaudatus* (Yi et al., 2019).

Large-scale genotyping at the whole genome level is becoming more and more important for population genetic studies (Zhang et al., 2019). Herein, we collected 8 *T. tenuis* populations and designed this study by the GBS method with the following objectives: 1) investigate fine-scale genetic variations of *T. tenuis* populations by the GBS database; 2) provide genomic evidence on the population structure of *T. tenuis* by a phylogenetic tree, admixture analysis, and principal component analysis; 3) determine the genetic diversity of 8 *T. tenuis* populations by detecting the genetic parameters, as well as explore the population genetic differentiation of 8 *T. tenuis* populations by the pairwise F_{ST} and genetic distance. This study is essential to the germplasm resource conservation of *T. tenuis* in Xinjiang Tarim River, as well as provides a valuable reference for genome-assisted breeding of *T. tenuis*.

MATERIALS AND METHODS

Samples Information

In this study, a total of 120 *T. tenuis* samples were collected from the branches of the Tarim River, including the Krakech River (WLWT, n = 15), the Yurungkash River (LF, n = 15), the Yarkant River (AETS, n = 15), the Kashgar River (SF, n = 15), the Toxkan River (WS, n = 15), the Muzart River (DWQ, n = 15; TKX, n = 15; KZE, n = 15) in 2021 (**Figure 1**). The fin tissues of all samples were collected and preserved in 95% ethanol. Total DNA was extracted using a DNA extraction kit (Tiangen, Beijing). The quality and quantity of DNA were detected by NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, United States), agarose gel electrophoresis and Agilent 2,100 Bioanalyzer (Agilent, Santa Clara, CA, United States).

Library Construction and Quality Control

The library was constructed by Super-GBS method (Qi et al., 2018) as follows: DNA was digested by PstI HF/MspI; T4 ligase was used to add linker and barcode at both ends of the digested fragment; 300–700 BP fragments were recovered by adjusting the volume ratio of magnetic bead solution to the connecting product; The recovered fragments were amplified by PCR with high fidelity enzyme. The mixed library was sequenced by Illumina NovaSeq platform with 150 paired-end. We used stacks (Catchen et al., 2013) software and split the offline data according to barcode and enzyme digestion site information to obtain raw reads of each sample. Moreover, the raw reads were filtered with the fastp program (Chen et al., 2018), and the quality filtering standards were as follows: remove the joint sequence; remove reads with N (non AGCT) base greater than or equal to 5; remove reads with an average base mass value less than 20.

Genome Construction and Variation Detection

The ustacks program in Stacks software (Catchen et al., 2013) was used to cluster the sequenced reads of each sample, while cstacks, sstacks, tsv2bam, and gstacks programs were used for genome construction and variation detection. In addition, we used vcftools software (Danecek et al., 2011) to filter the SNP typing results. The filtering conditions were as follows: non-second alleles were excluded; The number of reads support (DP) was not less than 4; the loci with MAF less than 0.01 were excluded; the loci with SNP typing deletion rate higher than 20% were excluded.

Population Genetic Structure Analysis

To explore the population structure of *T. tenuis* with different geographical distribution, admixture software (v1.3.0) (Alexander et al., 2009) was used to cluster all samples with a bayesian approach. The *K* value was determined according to ten-fold cross-validation. Pong software (Behr et al., 2016) was used to cluster the repeated results of each *K* value. The phylogenetic tree was constructed by the neighbor-joining (NJ) method, the distance matrix was calculated by treebest software (Ruan et al., 2008). The gcta software (Yang et al., 2011) was used to perform principal component analysis (PCA) by genome-wide SNPs information, and the scatter plotting was performed with the first three principal components using the "ggplot2" package in R (Ginestet, 2011).

Genetic Differentiation and Genetic Diversity Analyses

The genetic differentiation coefficient (F_{ST}) between populations was calculated using the R package genepop (Rousset, 2008). The F_{ST} values ranging from 0 to 0.05, 0.05 to 0.15, and 0.15 to 0.25 indicate that there are no genetic, moderate, and large differentiations among populations, respectively (Wright, 1965). The genetic distance (DR) between populations was estimated by F_{ST} value (Reynolds et al., 1983). Furthermore, several genetic indicators were used to investigate the genetic diversity of 8 *T. tenuis* populations, including Hardy-Weinberg equilibrium *p*-value (*HW-P*), Hardy-Weinberg (*He*), observed heterozygosity (*Ho*), polymorphism information content (*PIC*), effective number of alleles (*Ne*), observed number of alleles (*Na*), nucleotide diversity (*Pi*). The vcftools software was used to calculate these genetic indicators (Danecek et al., 2011).

RESULTS

Identification and Screening of SNP Markers

In this study, a total of 120 *T. tenuis* samples were sequenced and genotyped using GBS method. The constructed reference genome contains 572,808 sequences. After filtering out the raw reads, the total clean reads for all the genotypes were 582,678,756 with the average reads per sample being 4,855,656 (**Supplementary Table S1**). The average clean reads percent for all samples was 96.88%, and the lowest and highest number of reads was 3,240,046 and 7,040,766, respectively. The content of clean bases ranged from 0.47 G to 1.03 G, the average GC content was 44.08%, and the percent of clean base varied from 92.82 to 97.82%. After processing the raw reads via the GBS pipeline, we obtained a total of 595,379 SNPs using the VCF filtering control thresholds for further genetic analysis. The average sequencing depth of all samples was 16.28 X. Clean reads of each sample demonstrated high Q20 (>95.68%) and Q30 (>89.24%), indicating the high quality of sequencing data.

Population Genetic Structure Analysis of *T. tenuis* Populations

To investigate the grouping of 8 *T. tenuis* populations, we constructed the NJ tree of the 120 individuals, which indicated that different geographical populations of *T. tenuis* could be divided into five branches (**Figure 2A**). The individuals from WS and SF were clustered into cluster 1. The individuals from WLWT and LF were grouped into one subgroup and clustered into cluster 2 with the AETS individuals. Additionally, the KZE, DWQ, and TKX individuals were tightly clustered together to form cluster 3 (**Figure 2B**). The scatter plots of PCA showed that the WS individuals were separate from the SF individuals and that the DWQ, KZE and TKX individuals are close to each other, but distinct from AETS, WLWT and LF individuals (**Supplementary Figure S1**). Pong analysis was used to perform the genetic clusters for all individuals (K = 2–10), and nine independent runs for each K value were conducted (**Figure 3A**). Admixture analysis indicated that the minimizing K value was 3 (**Figure 3B**), suggesting that all *T. tenuis* individuals could be divided into three subgroups, which was consistent with the result of the phylogenetic tree. Therefore, it was revealed that the individuals from SF and WS showed different ancestry information compared with the individuals from the DWQ, KZE and TKX, and the individuals from the AETS, LF and WLWT.

Genetic Differentiation Among *T. tenuis* Populations

The genetic differentiation among the *T. tenuis* populations was identified by pairwise F_{ST} values and genetic distance (**Figure 4**). The

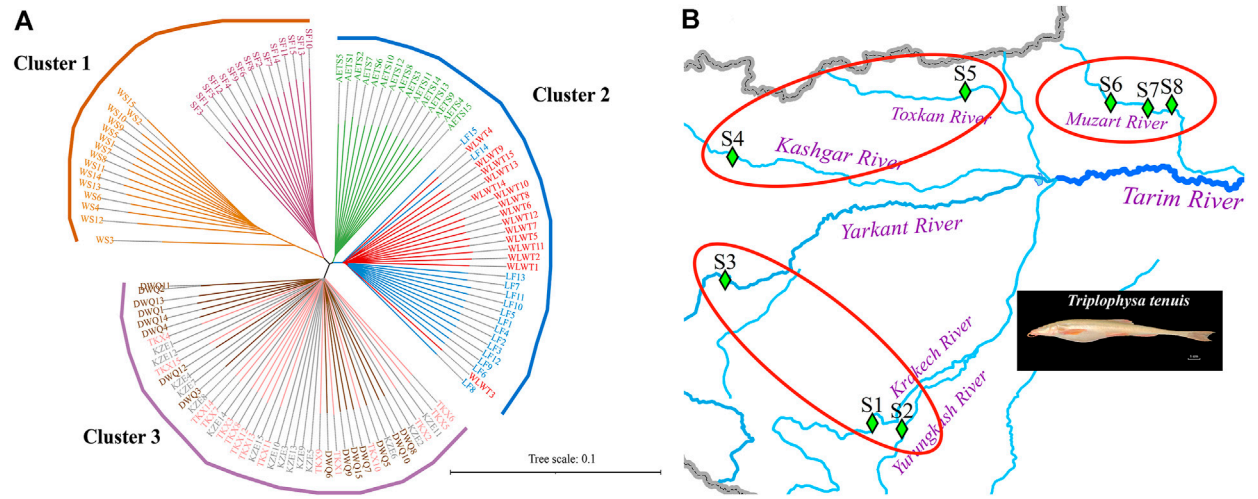


FIGURE 2 | Phylogenetic tree analysis of 8 *T. tenuis* populations. **(A)** Neighbor-Joining tree based on SNP data of 120 *T. tenuis* individuals. **(B)** Phylogenetic tree results corresponding to geographical location. The blue lines represent the rivers. The green square symbols indicate the sampling locations. S1, WLWF; S2, LF; S3, AETS; S4, SF; S5, WS; S6, DWQ; S7, TKX; S8, KZE.

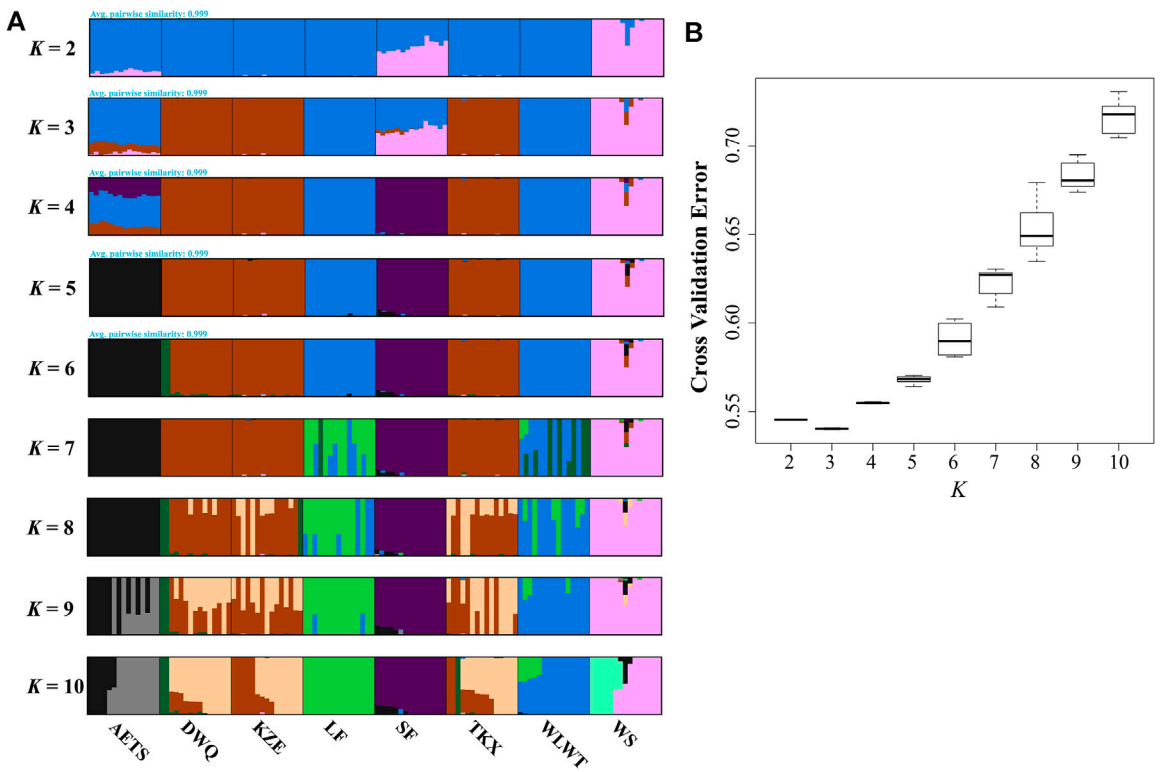


FIGURE 3 | Population genetic structure based on SNP data of 120 *T. tenuis* individuals. **(A)** Clustering information for different populations when $K = 2$–10; **(B)** The CV error varies among K values.

F_{ST} values between pairs of the 8 *T. tenuis* populations varied from 0.0010 (KZE with TKX) to 0.2224 (WLWT with WS), with an overall F_{ST} value (0.1126), suggesting a moderate genetic differentiation among these *T. tenuis* populations. Notably, we found that the F_{ST} values ranged from 0.1777 (WS with SF) to 0.2224 (WS with WLWT), indicating that there was a large genetic differentiation between WS population and the other *T. tenuis* populations. Moreover, the genetic distance between pairs of the

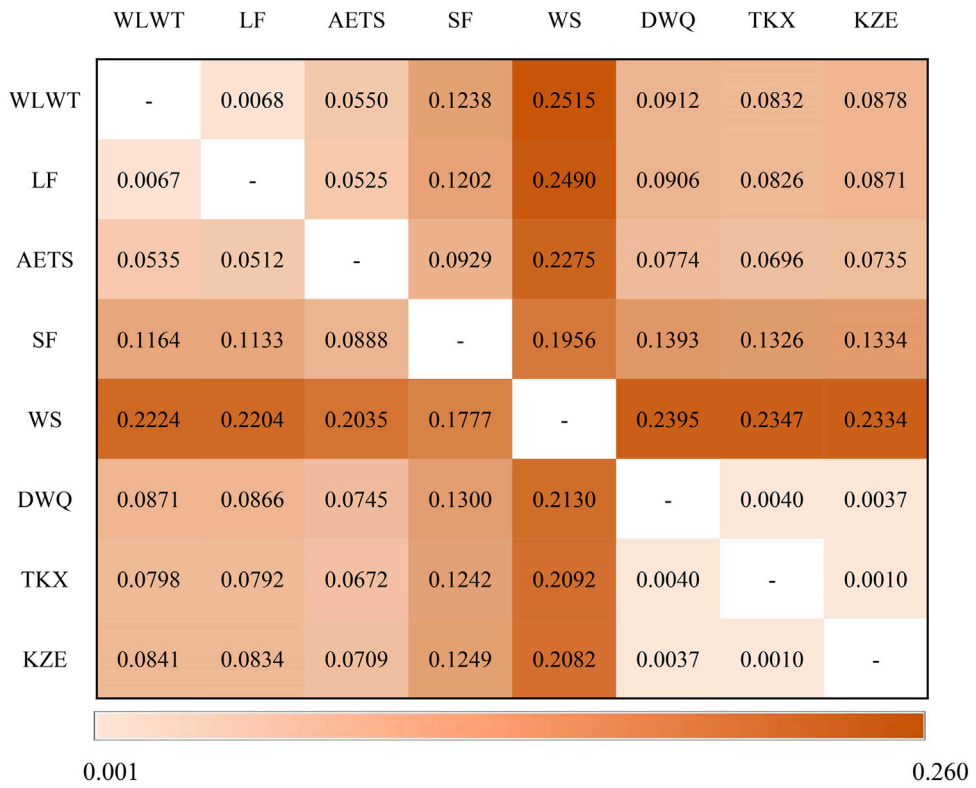


FIGURE 4 | Genetic differentiation F_{ST} value (below diagonal) and genetic distance (above diagonal).

8 *T. tenuis* populations ranged from 0.0010 (KZE with TKX) to 0.2515 (WLWT with WS). Likewise, there was a large genetic distance value between WS population and the other 7 *T. tenuis* populations, indicating an obvious genetic differentiation. By these analyses, we observed that there was a large genetic differentiation between WS and WLWT, a low genetic differentiation between KZE and TKX.

Genetic Diversity of *T. tenuis* Populations

The genetic parameters are shown in **Supplementary Table S2**. The *HW-P* index showed the relationship between allele frequency and genotype frequency. We found that *HW-P* values ranged from 0.8495 (WS) to 0.8777 (WLWT), indicating that all populations reached genetic equilibrium. The *He* index represents the expected value of heterozygosity based on Hardy Weinberg equilibrium, which depends only on allele frequency. The result showed that *He* values ranged from 0.1298 (WLWT) to 0.1450 (WS), with an average of 0.1359. The *Ho* index indicates the proportion of heterozygotes observed in a population. We observed that *Ho* values ranged from 0.1198 (WLWT) to 0.1307 (SF), with an average of 0.1245. The *PIC* is an indicator of polymorphism level. The result showed that *PIC* values ranged from 0.1085 (WLWT) to 0.1214 (WS), with an average of 0.1139. Additionally, *Ne*, *Na* and *Pi* indices are the basic parameters of genetic diversity, the result indicated that the *Ne* values ranged from 1.2011 (WLWT) to 1.2268 (SF) with an average of 1.2100, and the *Na* values ranged from 1.5584 (WLWT) to 1.6172 (WS) with an average of 1.5931. Moreover, the *Pi* values ranged from 0.1348 (WLWT) to 0.1508 (WS), with an average of 0.1411.

DISCUSSION

The harsh environment, characterized by high altitude, cold habitat and poor nutrition in the Qinghai-Tibet Plateau, has brought severe challenges to local species (Yang et al., 2015). Nevertheless, as the representative indigenous species in the Qinghai-Tibet Plateau, *Triplophysa* fishes can well adapt to these severe natural conditions (Chen et al., 2020). Given their taxonomic status, evolutionary process, geographical distribution and biological characteristics, *Triplophysa* genus offers an attractive study model for the fish phylogeny, geological change, life evolution, and extreme environmental adaptation (Chen et al., 2020; Wu et al., 2020). In recent years, due to water diversion, irrigation and climatic change, coupling factors, including the lower reach dry-out, river salinization, habitat fragmentation, etc., drove *T. tenuis* to only inhabit the upper reaches of this river, and caused its population decline (Yaning et al., 2009). To investigate the genetic diversity and genetic structure of *T. tenuis* populations, we collected the *T. tenuis* individuals from five sampling points (SF, WS, DWQ, TKX and KZE) in the north of Tarim River, and three sampling points (WLWT, LF, and AETS) in the south of Tarim River. Through population genetic structure analysis, we found that the individuals from the south of Tarim River were clustered together, the individuals from the north of Tarim River

were divided into two subgroups according to the geographical isolation.

Previous studies have revealed that geographical distance, ecological or environmental differences can bring about genetic isolation of populations and reduce the rate of successful migration (Chen and Wang, 2021; Gai et al., 2021). We found that the WS (Toxkan River) and SF individuals (Kashgar River) differentiated obviously and formed two separate clusters in cluster 1, and the AETS individuals (Yarkant River) differentiated with WLWT (Krakech River) and LF individuals (Yurungkash River) in cluster 2. The sampling points of Krakech River and Yurungkash River are relatively close in location, the individuals from WLWT showed a closer affinity to the LF. Additionally, the overall F_{ST} values indicated that there was a moderate gene differentiation among 8 *T. tenuis* populations, indicating that the genetic exchange among these *T. tenuis* populations was very restricted. Likewise, very large pairwise F_{ST} values were also estimated among *T. yarkandensis* populations in the Tarim River (Zhou et al., 2021). Tarim River is a typical seasonal and high salinity river, with an extremely fragile ecosystem and underdeveloped aquaculture. Thus, it is limited to mediated the genetic exchange among most populations mediated by agriculture activity. In the past decades, due to the influence of natural and anthropic factors, many branch rivers have successively lost surface water contact with the mainstream of the Tarim River. Therefore, it is inferred that the fragmentation of the Tarim River and its branch basins may be one of the important reasons for hindering the gene exchange among different *T. tenuis* populations.

Tarim River is the largest inland river in China, formed by the confluence of 114 rivers of 9 major water systems (Chen et al., 2007). The Krakech River is one of the main tributaries of the Hotan River, and has serious evaporation and seepage due to the middle reaches of this river crossing the Taklimakan Desert. Tashkurgan River is one of the main tributaries of the Yarkand River, but most of the water in the Yarkand River has been introduced into the reservoir since the Xiaohaizi Reservoir was constructed. Hence the two rivers only inject water into the mainstream of the Tarim River during the flood period. At present, Aksu River is the main water source of the Tarim River and supplies water to the Tarim River all year round (Hartmann et al., 2016). Herein, we collected the WS individuals from Toxkan River, which originates from the South Tianshan Mountains on the border between China and Kyrgyzstan, and merges with the Kumarak River to form the Aksu River. According to the pairwise F_{ST} values and genetic distance, we found that WS population had a large genetic differentiation with other *T. tenuis* populations. Moreover, it was indicated that WS population presented a higher genetic diversity. Therefore, it is suggested that the abundant water and the suitable habitat environment of the Tarim River have contributed to the formation of the unique genetic characteristics of *T. tenuis* (i.e., high genetic variation). Importantly, coupled with the fragmentation of the Tarim River, it may increase the possibility of outbreeding decline.

In conclusion, our study represented an important step in better understanding the genetic diversity and differentiation of *T. tenuis* within the Tarim River ecosystem, which is essential to develop the population management unit. Further, we envision that artificial propagation and releasing are conducted to restore the *T. tenuis* population in the Tarim River ecosystem. The genome-wide molecular markers can be available to establish a parent genetic information database and select an appropriate population for proliferation and release, which can not only be conducive to the restoration, but also avoid the genetic bottleneck effect and prevent genetic drift. Meanwhile, to maintain genetic diversity and avoid depletion of genetic resources, the genetic consequences of releasing the *T. tenuis* populations into the wild need to be evaluated.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://figshare.com/articles/dataset/Population_genetic_analysis_of_Triplophysa_tenuis/19126847/1.

ETHICS STATEMENT

The animal study was reviewed and approved by the Huazhong Agricultural University.

AUTHOR CONTRIBUTIONS

BH and XZ conceived the study. SC, XL, and JL extracted the DNA samples. BH, XL, XZ, QZ, JS, and RT collected the samples. BH and JC analyzed the data and wrote the draft of manuscript. JC, SC, and DL revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the Finance Special Fund of the Ministry of Agriculture and Rural Affairs "Fisheries Resources and Environment Survey in the Key Water Areas of Northwest China" (No. 2130111).

ACKNOWLEDGMENTS

We appreciate that students Jinfa Zhao, Kui Zhang, Lu Liu, Shilian Ding, Wenhao Zhao, Zhenyi Hu take the time to participate in the survey.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.860678/full#supplementary-material

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data. *Bioinformatics* 32, 2817–2823. doi:10.1093/bioinformatics/btw327

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an Analysis Tool Set for Population Genomics. *Mol. Ecol.* 22, 3124–3140. doi:10.1111/mec.12354

Chen, J., and Wang, W. (2021). Genetic Diversity and Genetic Differentiation of *Megalobrama* Populations Inferred by Mitochondrial Markers. *Genes Genom.* 43, 1119–1132. doi:10.1007/s13258-021-01126-8

Chen, Y.-n., Li, W.-h., Xu, C.-c., and Hao, X.-m. (2007). Effects of Climate Change on Water Resources in Tarim River Basin, Northwest China. *J. Environ. Sci.* 19, 488–493. doi:10.1016/s1001-0742(07)60082-5

Chen, Y., Ye, Z., and Shen, Y. (2011). Desiccation of the Tarim River, Xinjiang, China, and Mitigation Strategy. *Quat. Int.* 244, 264–271. doi:10.1016/j.quaint.2011.01.039

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Chen, S.-A., Hou, J., Yao, N., Xie, C., and Li, D. (2020). Comparative Transcriptome Analysis of *Triplophysa Yarkandensis* in Response to Salinity and Alkalinity Stress. *Comp. Biochem. Physiol. D: Genom. Proteomics* 33, 100629. doi:10.1016/j.cbd.2019.100629

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Gai, Z., Zhai, J., Chen, X., Jiao, P., Zhang, S., Sun, J., et al. (2021). Phylogeography Reveals Geographic and Environmental Factors Driving Genetic Differentiation of Populus Sect. Turanga in Northwest China. *Front. Plant Sci.* 12, 705083. doi:10.3389/fpls.2021.705083

Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc. A. Stat.* 174, 245–246. doi:10.1111/j.1467-985x.2010.00676_9.x

Hartmann, H., Snow, J. A., Su, B., and Jiang, T. (2016). Seasonal Predictions of Precipitation in the Aksu-Tarim River basin for Improved Water Resources Management. *Glob. Planet. Change* 147, 86–96. doi:10.1016/j.gloplacha.2016.10.018

Islam, A. S. M. F., Sanders, D., Mishra, A. K., and Joshi, V. (2021). Genetic Diversity and Population Structure Analysis of the USDA Olive Germplasm Using Genotyping-By-Sequencing (GBS). *Genes* 12, 2007. doi:10.3390/genes12122007

Katsumura, T., Oda, S., Mitani, H., and Oota, H. (2019). Medaka Population Genome Structure and Demographic History Described via Genotyping-By-Sequencing. *G3 Genes Genom. Genet.* 9, 217–228. doi:10.1534/g3.118.200779

Liu, M. C., Xu, X., Nie, Z. L., Wei, Z. T., Han, F., and Zhang, J. (2021). Embryonic Development of *Triplophysa Tenuis* from Kaidu River Basin. *J. Hydroecol.* 42, 72–78. (in Chinese, with abstract in English). doi:10.15928/j.1674-3075.201903160063

Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., et al. (2018). UGbS-Flex, a Novel Bioinformatics Pipeline for Imputation-free SNP Discovery in Polyploids without a Reference Genome: finger Millet as a Case Study. *BMC Plant Biol.* 18, 117. doi:10.1186/s12870-018-1316-3

Reynolds, J., Weir, B. S., and Cockerham, C. C. (1983). Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics* 105, 767–779. doi:10.1093/genetics/105.3.767

Rousset, F. (2008). genepop'007: a Complete Re-implementation of the Genepop Software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi:10.1111/j.1471-8286.2007.01931.x

Ruan, J., Li, H., Chen, Z., Coghlan, A., Coin, L. J., Guo, Y., et al. (2008). TreeFam: 2008 Update. *Nucleic Acids Res.* 36, D735–D740. doi:10.1093/nar/gkm1005

Tursun., Ren, B., Zhang, R.M., Ma, Y.W., and Guo, Y. (2005). The Study on the Biology of *Triplophysa* (*Triplophysa Teunis* (Day)) in Akesu River. *Chin. J. Fish.* 18, 65–69. (in Chinese, with abstract in English). doi:10.3969/j.issn.1005-3832.2005.02.012

Wang, N., Yuan, Y., Wang, H., Yu, D., Liu, Y., Zhang, A., et al. (2020). Applications of Genotyping-By-Sequencing (GBS) in maize Genetics and Breeding. *Sci. Rep.* 10, 16308. doi:10.1038/s41598-020-73321-8

Wright, S. (1965). The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* 19, 395–420. doi:10.1111/j.1558-5646.1965.tb01731.x

Wu, H., Gu, Q., Zhou, C., Tang, Y., Husemann, M., Meng, X., et al. (2020). Molecular Phylogeny and Biogeography of *Triplophysa Stone* Loaches in the Central Chinese Mountains. *Biol. J. Linn. Soc.* 130, 563–577. doi:10.1093/biolinnean/blaa060

Xie, C. G., Ma, Y. W., and Guo, Y. (2015). Analysis of Biogeography of Fishes in Tarim Basin. *Chin. J. Fish.* 28, 40–46. (in Chinese, with abstract in English). doi:10.1071/sb14023

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88, 76–82. doi:10.1016/j.ajhg.2010.11.011

Yang, Y., Wang, L., Han, J., Tang, X., Ma, M., Wang, K., et al. (2015). Comparative Transcriptomic Analysis Revealed Adaptation Mechanism of *Phrynocephalus Erythrurus*, the Highest Altitude Lizard Living in the Qinghai-Tibet Plateau. *BMC Evol. Biol.* 15, 101. doi:10.1186/s12862-015-0371-8

Yaning, C., Changchun, X., Xingming, H., Weihong, L., Yapeng, C., Chenggang, Z., et al. (2009). Fifty-year Climate Change and its Effect on Annual Runoff in the Tarim River Basin, China. *Quat. Int.* 208, 53–61. doi:10.1016/j.quaint.2008.11.011

Yi, S., Wang, W., and Zhou, X. (2019). Genomic Evidence for the Population Genetic Differentiation of *Misgurnus anguillicaudatus* in the Yangtze River basin of China. *Genomics* 111, 367–374. doi:10.1016/j.ygeno.2018.02.011

Younis, A., Ramzan, F., Ramzan, Y., Zulfiqar, F., Ahsan, M., and Lim, K. B. (2020). Molecular Markers Improve Abiotic Stress Tolerance in Crops: A Review. *Plants* 9, 1374. doi:10.3390/plants9101374

Zhang, B.-D., Xue, D.-X., Li, Y.-L., and Liu, J.-X. (2019). RAD Genotyping Reveals fine-scale Population Structure and Provides Evidence for Adaptive Divergence in a Commercially Important Fish from the Northwestern Pacific Ocean. *Peer J.* 7, e7242. doi:10.7717/peerj.7242

Zhang, B., Jia, L., He, X., Chen, C., Liu, H., Liu, K., et al. (2020). Large Scale SNP Unearthing and Genetic Architecture Analysis in Sea-Captured and Cultured Populations of Cynoglossus Semilaevis. *Genomics* 112, 3238–3246. doi:10.1016/j.ygeno.2020.06.013

Zhong, J., Yi, S., Ma, L., and Wang, W. (2019). Evolution and Phylogeography Analysis of Diploid and Polyploid *Misgurnus anguillicaudatus* Populations across China. *Proc. R. Soc. B.* 286, 20190076. doi:10.1098/rspb.2019.0076

Zhou, X., Yi, S., Zhao, W., Zhou, Q., Shen, J., Li, D., et al. (2021). Genetic Diversity and Population Differentiation of Kashgarian Loach (*Triplophysa Yarkandensis*) in Xinjiang Tarim River Basin. *Biology* 10, 734. doi:10.3390/biology10080734

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fevo.2022.783215

Population Genetics Reveals Invasion Origin of *Coilia brachygnathus* in the Three Gorges Reservoir of the Yangtze River, China

Dongdong Zhai[1], Bo Li[2,3], Fei Xiong[1], Wei Jiang[2,3]*, Hongyan Liu[1], Chenhao Luo[1], Xinbin Duan[4] and Daqing Chen[4]*

[1] Hubei Engineering Research Center for Protection and Utilization of Special Biological Resources in the Hanjiang River Basin, School of Life Sciences, Jianghan University, Wuhan, China, [2] Hubei Key Laboratory of Three Gorges Project for Conservation of Fishes, Yichang, China, [3] Chinese Sturgeon Research Institute, China Three Gorges Corporation, Yichang, China, [4] Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan, China

OPEN ACCESS

Edited by:
Narongrit Muangmai, Kasetsart University, Thailand

Reviewed by:
Antonia Picornell, University of the Balearic Islands, Spain
Kathirvelpandian Ayyathurai, National Bureau of Fish Genetic Resources (ICAR), India

***Correspondence:**
Fei Xiong
xf9603@163.com
Wei Jiang
jiang_wei6@ctg.com.cn

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *25 September 2021*
Accepted: *31 January 2022*
Published: *04 March 2022*

Citation:
Zhai D, Li B, Xiong F, Jiang W, Liu H, Luo C, Duan X and Chen D (2022) Population Genetics Reveals Invasion Origin of Coilia brachygnathus in the Three Gorges Reservoir of the Yangtze River, China. Front. Ecol. Evol. 10:783215. doi: 10.3389/fevo.2022.783215

Non-indigenous fish invasions have posed a serious threat to global fish diversity and aquatic ecosystem security. Studying the invasion sources, pathways, and genetic mechanisms by means of population genetics is helpful in the management and control of non-indigenous fishes. In this study, we used mitochondrial Cyt *b* gene, D-Loop region and microsatellite markers to analyze the genetic diversity and population structure of 12 *Coilia brachygnathus* populations from the native and invaded regions of the Yangtze River Basin in order to explore the invasion sources, pathways, and genetic mechanisms of *C. brachygnathus* in the Three Gorges Reservoir. The results showed that the main invasion sources of *C. brachygnathus* in the Three Gorges Reservoir were the Poyanghu Lake, Dongtinghu Lake, Changhu Lake, and other populations in the middle reaches of the Yangtze River. The invasion pathway may have involved moving upstream through the operation of ship locks. The genetic diversity of *C. brachygnathus* in the invasive populations was significantly smaller than in the native populations, indicating a founder effect. The low genetic diversity did not affect the successful invasion, confirming that genetic diversity and successful invasion do not always have a simple causal relationship. These results can provide basic data for the prevention and control of *C. brachygnathus* in the Three Gorges Reservoir and study case for understanding the mechanism of invasion genetics.

Keywords: *Coilia brachygnathus*, Three Gorges Reservoir, invasion sources, invasion pathways, invasion genetic mechanisms

INTRODUCTION

Biological invasion is a main concern to conservationists and resource managers owing to its potential threat to ecosystems and biodiversity (Mack et al., 2000; Dudgeon et al., 2006). Invaders may drastically change the functioning of the ecosystem where they are introduced, resulting in the decline or extinction of native species through competition, predation, and habitat alteration (Levine, 2008; Haubrock et al., 2021). Identifying the putative source populations and the invasion

pathways is a key first step in developing appropriate management measures to mitigate the impacts of species invasions (Bariche et al., 2017; Morim et al., 2019). Population genetic research is a powerful tool for reconstructing the invasion history (Estoup and Guillemaud, 2010).

The Yangtze River is the largest river in China, the upper reaches of the river are turbulent, and the middle and lower reaches of the river are open and the flow of water is slow. Yichang City is the dividing point between the upper reaches and the middle and lower reaches of the Yangtze River (Fish research laboratory of Institute of Hydrobiology of Hubei province, 1976). *Coilia brachygnathus* (Clupeiformes, Engraulidae, *Coilia*) is an anchovy species that prefer to live in still water environment (Whitehead et al., 1988). It was first described from the Dongtinghu Lake of the Yangtze River Basin (Kreyenberg and Pappenheim, 1908). Therefore, in the past, due to the influence of the river water velocity, it has been mainly distributed in the middle and lower reaches of the Yangtze River, especially the lakes in this section, such as Poyanghu Lake, Dongtinghu Lake, and Changhu Lake, and there was no record of the distribution of *C. brachygnathus* in the upper reaches of the Yangtze River (Fish research laboratory of Institute of Hydrobiology of Hubei province, 1976; Ding, 1994). However, *C. brachygnathus* accounted for 0.15% of the total fish assemblage biomass in the Three Gorges Reservoir of the upper Yangtze River according to investigate from 2015 to 2017 (Liao et al., 2018). Surprisingly, it has become an absolutely dominant species in the Three Gorges Reservoir according to our investigations from 2019 to 2021 (accounted for 18.90% of the total fish assemblage biomass in the Three Gorges Reservoir, unpublished data).

How did *C. brachygnathus* invade the Three Gorges reservoir? In 1981, the Gezhouba Dam was completed in the Yichang section of the upper reaches of the Yangtze River (Wei et al., 1997). Then in 2003, the Three Gorges Dam was completed 38 km upstream of the Gezhouba Dam (Xu and Milliman, 2009). The completion of these two dams, especially the construction of the Three Gorges Dam, the largest hydropower station in the world, has made the Gezhouba Reservoir and the Three Gorges Reservoir changed from turbulent water habitats to static water environments, this created feasible conditions for the survival of fish that prefer to live in still water environment, such as *C. brachygnathus* (Huang and Wu, 2018; Gao et al., 2019). (In the present study, we regard the Gezhouba Reservoir as part of the Three Gorges Reservoir; so, only the Three Gorges Reservoir will be mentioned hereafter). As for the way of *C. brachygnathus* invaded the Three Gorges Reservoir? We guess that it is most likely to spread to the Three Gorges Reservoir through the ship locks. For shipping, the Three Gorges Dam has built double-line five-grade ship locks (Stone, 2008). Similarly, Gezhouba Dam has also built three ship locks (Lin et al., 2013). While these locks provide convenience for shipping, they also create conditions for the proliferation of certain fish (Lin et al., 2013). Ships and fish are transported from below the dam to above the dam like climbing the stairs.

Moreover, how *C. brachygnathus* developed into the dominant species in the Three Gorges Reservoir so quickly, it may be related to factors such as sufficient food and strong ability to resist predation, as well as its high fecundity. It is a carnivorous fish that mainly feeds on zooplankton, small fish and river prawns (Fish research laboratory of Institute of Hydrobiology of Hubei province, 1976; Zhang et al., 2013). Due to overfishing of the middle and large sized fishes, many studies have showed that the small-sized fishes accounted for a great proportion in fish catches of the Three Gorges Reservoir in the past 10 years (Gao et al., 2010, 2019; Wei et al., 2021). As the water flow slowed down, the eutrophication degree of the Three Gorges Reservoir has increased, so the zooplankton has increased too (Li et al., 2019). *Coilia brachygnathus* has serrations on the edges of the upper and lower jaws and abdomen that help it to defense against predation (Meyers et al., 2008). The absolute fecundity of *C. brachygnathus* varies between 6,905 and 28,745 eggs with a mean value of 14,049, showing a strong reproductive ability (Liu, 2008).

Beginning in 2020, a "10-year ban on fishing" has been implemented in the Yangtze River (Liu et al., 2021). With the disappearance of fishing pressure, the population of *C. brachygnathus* in the Three Gorges Reservoir may further increase, and this will pose a serious threat to aquatic organisms and ecosystems in the Three Gorges Reservoir. It is necessary to identify the invasion sources and pathways of *C. brachygnathus* in the Three Gorges Reservoir for formulating management measures. Therefore, we used mitochondrial Cyt *b* gene, D-Loop region and microsatellite markers to analyze the population structure and genetic diversity of 12 *C. brachygnathus* populations from the donor and recipient regions in the Yangtze River Basin in order to gain insights into the invasion origin and process of the non-indigenous fish *C. brachygnathus* in the Three Gorges Reservoir. The results will provide a theoretical basis for formulating management measures to mitigate the impacts of *C. brachygnathus* invasion and understanding the mechanism of invasion genetics.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Samples of *C. brachygnathus* were collected from 12 different sampling sites in the Yangtze River Basin from August 2019 to August 2020 (Each sample site was more than 30 individual). From downstream to upstream, these were Taihu Lake (TH), Chaohu Lake (AHCH), Poyanghu Lake (PYH), Dongtinghu Lake (DTH), Changhu Lake (CH), Sandouping (SDP), TaipingXi (TPX), Xiangxi River (XXH), Xiaojiang River (XJ), Wanzhou (WZ), Fengdu (FD), and Fuling (FL). TH and AHCH are affiliated lakes in the lower reaches of the Yangtze River; PYH, DTH, and CH are affiliated lakes in the middle reaches of the Yangtze River, the original distribution of *C. brachygnathus* in the Yangtze River Basin. SDP is located between the Gezhouba Dam and the Three Gorges Dam; TPX, XXH, XJ, WZ, FD, and FL belong to the Three Gorges Reservoir, of which XXH and XJ are tributaries of the Three Gorges Reservoir, the invasion region of *C. brachygnathus* in the Yangtze River Basin (**Figure 1**). Through our investigations, it was found that *C. brachygnathus* was scarce in the river section above Fuling in the Three Gorges Reservoir, where is still flowing water

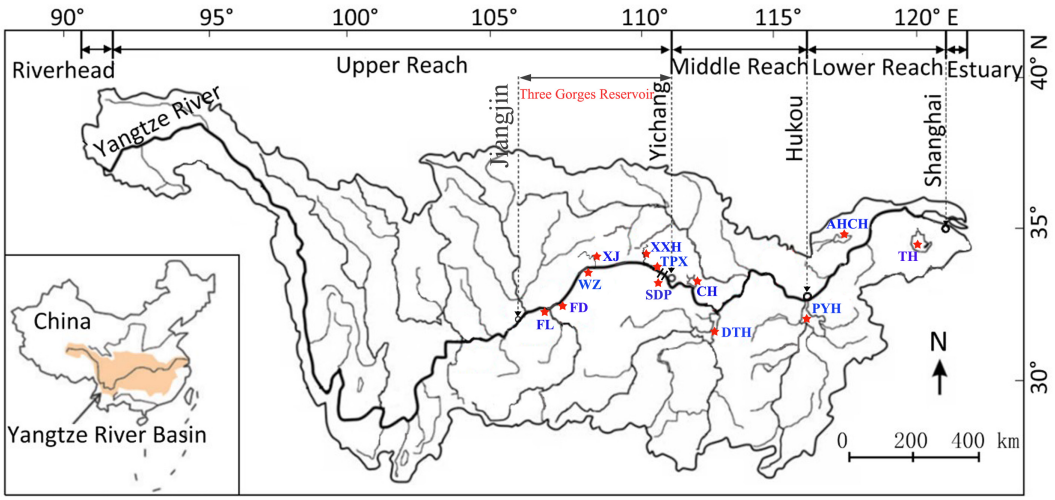


FIGURE 1 | The 12 sampling sites of *C. brachygnathus* in the Yangtze River Basin. TH, Taihu Lake; AHCH, Chaohu Lake; PYH, Poyanghu Lake; DTH, Dongtinghu Lake; CH, Changhu Lake; SDP, Sandouping; TPX, TaipingXi; XXH, Xiangxi River; XJ, Xiaojiang River; WZ, Wanzhou; FD, Fengdu; FL, Fuling.

habitat. Therefore, these 12 sampling sites basically covered the distribution range of *C. brachygnathus* in the Yangtze River Basin. Species identification mainly refers to Fauna Sinica Osteichthyes Cypriniformes II (Chen, 1998).

After the fish samples were collected, tissues from the dorsal muscles were cut and stored in 95% alcohol at −20°C in a centrifuge tube. Approximately 50 mg of muscles were used to extract genomic DNA using the animal tissue genomic DNA extraction kit of Chengdu Forge Biotechnology Co., Ltd.

mtDNA Amplification and Sequencing

The amplification primers for the mitochondrial DNA Cyt *b* fragment were L14724 5′-GAC TTG AAA AAC CAC CGT TG-3′ (forward) and H15915 5′-CTC CGA TCT CCG GAT TAC AAG AC-3′ (reverse) (Xiao et al., 2001). The amplification primers for the mitochondrial DNA D-Loop fragment were DF1 5′-CTA ACT CCC AAA GCT AGA ATT CT-3′ (forward) and DR2 5′-ATC TTA GCA TCT TCA GTG-3′ (reverse) (Tang et al., 2007). The total volume for the Polymerase Chain Reaction (PCR) was 30 μL: 15 μL of 2 × Taq PCR MasterMix (containing Taq DNA polymerase, dNTPs, $MgCl_2$, reaction buffer, Beijing Adler Biotechnology Co., Ltd.), template DNA 3 μL, forward and reverse primers 1 μL each, ddH_2O 10 μL. The PCR amplification procedure was as follows: pre-denaturation at 94°C for 4 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 45 s, extension at 72°C for 1 min, and finally extension at 72°C for 10 min. After the PCR products were subjected to agarose gel electrophoresis, the samples with accurate and clear target bands were sent to Tianyi Huiyuan Biotechnology Co., Ltd. for purification, recovery, and sequence determination. The mitochondrial DNA Cyt *b* fragment was measured in both directions, and the D-Loop fragment was only sequenced in the reverse direction due to the presence of repeated fragments in the forward direction.

Simple Sequence Repeats Amplification and Genotyping

There are no published microsatellite [also known as simple sequence repeats (SSR)] primers for *C. brachygnathus*, but many microsatellite primers have been published for *Coilia nasus*. Because *C. brachygnathus* and *C. nasus* are closely related species, a total of 50 pairs of microsatellite primers of *C. nasus* were selected from references for primer screening with eight DNA templates of *C. brachygnathus* (Du et al., 2019; Yu et al., 2019). Finally, eight pairs of primers with higher polymorphism were screened out by the capillary electrophoresis method. The specific sequences, repeat motif, and optimum annealing temperature of each primer are listed in **Supplementary Table 1**. Then, the 5′ ends of the eight pairs of forward primers were labeled with fluorescent groups (FAM). The primer screening and fluorescent labeling of the primers were entrusted to Tianyi Huiyuan Biotechnology Co., Ltd. The total volume for the PCR was 10 μL: 5 μL of 2 × Taq PCR MasterMix (containing Taq DNA polymerase, dNTPs, $MgCl_2$, reaction buffer, Beijing Adler Biotechnology Co., Ltd.), template DNA 1 μL, forward and reverse primers 0.5 μL each, ddH_2O 3 μL. The PCR amplification procedure was as follows: pre-denaturation at 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at suitable temperature for 40 s, extension at 72°C for 1 min, and finally extension at 72°C for 10 min. The amplified products were sent to Tianyi Huiyuan Biotechnology Co., Ltd. for microsatellite genotyping.

mtDNA Sequence Analysis

MEGA7 was used to align and rectify the nucleotide sequences by referring to the sequencing map (Kumar et al., 2016). DnaSP v5.10 was used to calculate the numbers of variable sites and haplotypes, haplotype diversity (H_d), nucleotide diversity (P_i), and gene flow (N_m) between groups (Librado and Rozas, 2009). *T*-test was used to test whether the genetic diversity

parameters or gene flow of different groups are significantly different. Arlequin version 3.0 was used to perform analysis of molecular variance (AMOVA) and to compute pairwise genetic differentiation index F_{ST} values (Excoffier et al., 2005). MEGA7 was used to construct genetic distance trees by the neighbor-joining method (Kumar et al., 2016). Network 4.6 was used to construct haplotype network using the median-joining algorithm (Bandelt et al., 1999).

Simple Sequence Repeats Data Analysis

Genotypes were checked for large allele dropout, null alleles, and scoring errors due to stuttering by Micro-checker v2.2.1 (Van Oosterhout et al., 2004). Tests for deviations from Hardy-Weinberg Equilibrium (HWE) and Linkage Disequilibrium (LD) across all pairs of loci were checked by GENEPOP v4.7.0 (Rousset, 2008) using exact tests with a Markov chain algorithm (*P*-values were estimated from 10,000 dememorizations, 100 batches, and 5,000 iterations per batch) (Guo and Thompson, 1992). Significance levels were adjusted for multiple comparisons using the sequential Bonferroni correction (Rice, 1989).

Number of alleles (*A*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and polymorphic information content (PIC) per locus were calculated by Cervus v3.0 (Kalinowski et al., 2007). Standardized allelic richness (*Ar*) was calculated using Fstat v2.9.3.2 (Goudet, 2001). *T*-test was used to test whether the genetic diversity parameters or gene flow of different groups are significantly different. Arlequin v3.0 was used to perform AMOVA and to compute pairwise genetic differentiation index F_{ST} and gene flow index N_m (Excoffier et al., 2005). In addition, genetic differentiation among populations was depicted by two-dimensional plots from a principal components analysis (PCA) of the allele frequencies matrix in GENALEX 6.4 (Peakall and Smouse, 2006).

The number of genetically differentiated clusters (*K*) was inferred using Bayesian assignment analysis by STRUCTURE v2.3.4 (Pritchard et al., 2000). Each assumed *K* (1–12) were performed 10 replications under an admixture model and correlated allele frequencies within populations (400,000 iterations with 100,000 burn-in periods) (Falush et al., 2003). The optimal *K* was selected according to the mean log probability LnP (*K*) and the ΔK value for each *K* (Evanno et al., 2005) using Structure Harvester (Earl and vonHoldt, 2012).[1]

RESULTS

mtDNA Marker

Genetic Diversity

For the Cyt *b* gene, a 1 141 bp sequence was obtained for 424 individuals after alignment. No insertions or deletions were observed. Within the 1 141 bp region, 59 sites were variable, including 30 singleton variable sites and 29 parsimony informative sites. We identified 60 haplotypes (GenBank accession numbers: OK172180–OK172239) from the 424 individuals, and the numbers of haplotypes ranged from 2 to 23 for each sampled population. For the 12 populations, haplotype diversity ranged from 0.157 to 0.958, with a mean value of 0.643. Nucleotide diversity ranged from 0.051 to 0.347%, with a mean value of 0.278%. On the whole, whether concerning the number of haplotypes, haplotype diversity, or nucleotide diversity, the non-invaded habitat (TH, AHCH, PYH, DTH, CH) was significantly ($P < 0.05$) greater than the invaded habitat (SDP, TPX, XXH, XJ, WZ, FD, FL) (**Table 1**).

For the D-Loop region, an 882 bp sequence was obtained for 408 individuals after alignment. Within the 882 bp region, 63 sites were variable, including 27 singleton variable sites and 36 parsimony informative sites. We identified 78 haplotypes (GenBank accession numbers: OK172240–OK172317) from the 408 individuals, and the numbers of haplotypes ranged from 2 to 20 for each sampled population. For the 12 populations, haplotype diversity ranged from 0.114 to 0.946, with a mean value of 0.704. Nucleotide diversity ranged from 0.026 to 0.674%, with a mean value of 0.518%. On the whole, whether concerning the number of haplotypes, haplotype diversity, or nucleotide diversity, the non-invaded habitat (TH, AHCH, PYH, DTH, CH) was significantly ($P < 0.05$) greater than the invaded habitat (SDP, TPX, XXH, XJ, WZ, FD, FL) (**Table 1**).

Population Structure

For the Cyt *b* gene, the pairwise F_{ST} values suggested no significant genetic differentiation between TH and AHCH populations, but these were significantly different from all other populations. The three populations that represent the lakes attached to the middle reaches of the Yangtze River (PYH, DTH and CH) had no significant genetic differentiation, but there was significant genetic differentiation between them and some populations in the Three Gorges Reservoir. The seven populations in the Three Gorges Reservoir (SDP, TPX, XXH, XJ, WZ, FD, and FL) had no significant genetic differentiation. For the D-Loop region, the pairwise F_{ST} values suggested no significant genetic differentiation among the populations that represented the lakes attached to the middle reaches of the Yangtze River (PYH, DTH, and CH), and this was the same for the seven populations in the Three Gorges Reservoir (SDP, TPX, XXH, XJ, WZ, FD, and FL), while there was significant genetic differentiation between any other two populations (**Table 2**). Based on the results of the genetic analyses of the two mtDNA genes and geographic distance, we divided the 12 populations into three groups for AMOVA. The three groups were the lower Yangtze River (TH, AHCH), the middle Yangtze River (PYH, DTH, CH), and the Three Gorges Reservoir (SDP, TPX, XXH, XJ, WZ, FD, FL). The results suggested that most of the genetic variation occurred among groups and within populations, with only slight genetic variation (Cyt *b*: 0.51%; D-Loop: 0.61%) occurring among populations within groups, thus suggesting that the group division was reasonable (**Table 3**).

The results showed that the genetic distance tree could be divided into two branches, one being TH and AHCH and the other comprising PYH, DTH, CH, SDP, TPX, XXH, XJ, WZ, FD, and FL. This suggested that the Three Gorges Reservoir population is closer to the midstream population and farther from the downstream population (**Figure 2**). Consistent with

[1] http://taylor0.biology.ucla.edu/structureHarvester/

TABLE 1 | Genetic diversity of *C. brachygnathus* from 12 sample sites in the Yangtze River Basin.

Molecular marker	Parameter	TH	AHCH	PYH	DTH	CH	SDP	TPX	XXH	XJ	WZ	FD	FL	Total
Cyt *b*	Sample size	36	35	37	36	36	39	35	35	29	36	36	34	424
	Variable Site	26	28	18	18	19	4	4	4	3	4	4	4	59
	N	20	23	15	12	13	2	2	2	2	2	2	2	60
	H_d	0.930	0.958	0.818	0.689	0.829	0.335	0.420	0.393	0.192	0.157	0.286	0.258	0.643
	P_i	0.00247	0.00234	0.00242	0.00214	0.00347	0.00117	0.00147	0.00138	0.00051	0.00055	0.00100	0.00091	0.00278
D-Loop	Sample size	31	34	35	35	36	34	34	36	33	34	30	36	408
	Variable Site	38	30	26	24	26	2	3	3	2	2	2	2	63
	N	20	19	19	20	19	2	3	3	2	2	2	2	78
	H_d	0.940	0.888	0.929	0.946	0.914	0.258	0.415	0.427	0.170	0.114	0.287	0.286	0.704
	P_i	0.00630	0.00372	0.00599	0.00501	0.00674	0.00059	0.00091	0.00094	0.00039	0.00026	0.00065	0.00065	0.00518
SSR	Sample size	32	32	32	32	32	32	32	32	32	32	32	32	384
	A	7.375	8.125	5.750	5.750	5.125	3.500	3.500	3.625	3.375	3.250	3.250	3.125	10.875
	Ar	7.036	7.839	5.501	5.508	5.005	3.445	3.358	3.498	3.249	3.210	3.125	3.043	6.641
	Ho	0.531	0.597	0.637	0.619	0.627	0.461	0.482	0.561	0.445	0.485	0.512	0.530	0.541
	He	0.636	0.696	0.615	0.605	0.596	0.437	0.442	0.472	0.419	0.461	0.465	0.477	0.638

N, Number of haplotypes; H_d, Haplotype diversity; P_i, Nucleotide diversity; A, number of alleles; Ar, allelic richness; Ho, observed heterozygosity; He, expected heterozygosity.

this, the haplotype network indicated that there were two clades of all the haplotypes for both the Cyt *b* gene and D-Loop region. One clade was almost exclusively occupied by individuals from the TH and AHCH populations, with only a few individuals from the PYH, DTH, and CH populations, the other clade was occupied by individuals from the PYH, DTH, CH, SDP, TPX, XXH, XJ, WZ, FD, and FL populations, indicating that the invaded populations only shared haplotypes with midstream populations (**Figure 3**). The gene flow between the Three Gorges Reservoir and the middle Yangtze River was significantly ($P < 0.05$) greater than those between the Three Gorges Reservoir and the lower Yangtze River (**Table 4**).

Simple Sequence Repeats Marker

Genetic Diversity

A total of 384 samples from 12 localities and 8 polymorphic SSR loci were used in the present study. No evidence for scoring errors due to stuttering, large allele dropout, or null alleles was detected by Micro-Checker. Ten of 96 population × locus combinations were significantly different from the HWE after Bonferroni correction, whereas no significant deviation from the HWE was detected in any locus across all populations. No loci pairs showed evidence of LD after Bonferroni correction. When data from all populations were pooled, the number of alleles at each locus ranged from 6 to 25, so all the eight loci were polymorphic (**Supplementary Table 2**).

The observed allele, allelic richness, observed heterozygosity, and expected heterozygosity of each locus for the 12 *C. brachygnathus* populations were listed in **Supplementary Table 2**. The average number of alleles per population ranged from 3.125 to 8.125, and the average allelic richness per population ranged from 3.043 to 7.839. The average observed heterozygosity per population ranged from 0.445 to 0.637, whereas the average expected heterozygosity per population ranged from 0.419 to 0.696. On the whole, whether considering the number of alleles, allele richness, observed heterozygosity, or expected heterozygosity, the non-invaded habitat (TH, AHCH, PYH, DTH, CH) was significantly ($P < 0.05$) greater than the invaded habitat (SDP, TPX, XXH, XJ, WZ, FD, FL) (**Table 1**).

Population Structure

The pairwise F_{ST} values suggested no significant genetic differentiation among the three populations that comprised the lakes attached to the middle reaches of the Yangtze River (PYH, DTH, and CH), and the same was true for the six out of the seven invasive populations in the Three Gorges Reservoir (TPX, XXH, XJ, WZ, FD, and FL), but there was significant genetic differentiation between any other two populations (**Table 5**). We also divided the 12 populations into three groups for AMOVA; the results indicated that the sources of genetic variation were mainly among groups (25.71%) and within populations (73.12%); only a slight amount of genetic variation (1.17%) occurred among populations within groups (**Table 3**). When conducting STRUCTURE analysis, LnP(K) showed a peak when $K = 5$, inferring there were five genetic clusters in the *C. brachygnathus* samples (**Figure 4**). In the figure, the five colors represent five clusters. The results showed that TH and AHCH populations in

TABLE 2 | Pairwise F_{ST} values based on Cyt *b* analysis of 424 samples (below diagonal) and D-Loop analysis of 408 samples (above diagonal) of *C. brachygnathus* from 12 different sites in the Yangtze River Basin, China.

	TH	AHCH	PYH	DTH	CH	SDP	TPX	XXH	XJ	WZ	FD	FL
TH		0.04589*	0.54150*	0.57060*	0.46001*	0.70321*	0.69498*	0.69956*	0.70586*	0.71277*	0.68967*	0.70705*
AHCH	0.02870		0.61251*	0.64341*	0.52734*	0.79399*	0.78389*	0.78627*	0.79878*	0.80544*	0.78419*	0.79554*
PYH	0.58526*	0.60556*		0.00166	0.02402	0.21729*	0.18176*	0.18112*	0.24012*	0.25916*	0.19860*	0.21396*
DTH	0.61391*	0.63347*	−0.00112		0.02629	0.23924*	0.19849*	0.19759*	0.26537*	0.28657*	0.21932*	0.23488*
CH	0.48595*	0.50538*	0.00399	0.01187		0.16712*	0.13796*	0.13779*	0.18663*	0.20328*	0.15103*	0.16449*
SDP	0.68262*	0.70161*	0.06294	0.03267	0.08594*		−0.00448	0.00386	−0.01544	0.01223	−0.03088	−0.02794
TPX	0.66792*	0.68683*	0.04758	0.02650	0.06971	−0.00989		−0.02675	0.04239	0.08541	−0.01685	−0.01371
XXH	0.67044*	0.68953*	0.04915	0.02482	0.07186*	−0.02003	−0.02729		0.05425	0.09839	−0.01009	−0.00686
XJ	0.70162*	0.72289*	0.10767*	0.07163	0.11734*	0.04024	0.10862	0.08296		−0.02312	−0.00647	−0.00418
WZ	0.71349*	0.73348*	0.13446	0.09801*	0.14197*	0.03169	0.10287	0.07603	−0.01797		0.02745	0.02805
FD	0.68734*	0.70686*	0.07622	0.04314	0.09593	−0.02245	0.01195	−0.00402	0.01332	0.00357		−0.03152
FL	0.68933*	0.70919*	0.08465	0.05067	0.10181	−0.01653	0.02673	0.00802	0.00123	−0.00899	−0.02794	

**P < 0.05 after Bonferroni correction.*

TABLE 3 | Analysis of molecular variance (AMOVA) for 12 *C. brachygnathus* populations based on mtDNA and SSR analyses.

Molecular marker	Source of variation	*df*	Sum of squares	Variance components	Percentage variation	Fixation indices	*P*
Cyt *b*	Among groups	2	265.995	1.08072	52.82	0.52819	0
	Among populations within groups	9	11.882	0.01037	0.51	0.01074	0.07527
	Within populations	412	393.453	0.95498	46.67	0.53326	0
	Total	423	671.330	2.04607			
D-Loop	Among groups	2	440.146	1.87751	60.05	0.60046	0
	Among populations within groups	9	16.870	0.01894	0.61	0.01516	0.00782
	Within populations	396	487.215	1.23034	39.35	0.60652	0
	Total	407	944.230	3.12679			
SSR	Among groups	2	326.304	0.72696	25.71	0.25710	0
	Among populations within groups	9	37.704	0.03316	1.17	0.01578	0
	Within populations	756	1562.969	2.06742	73.12	0.26883	0
	Total	767	1926.977	2.82754			

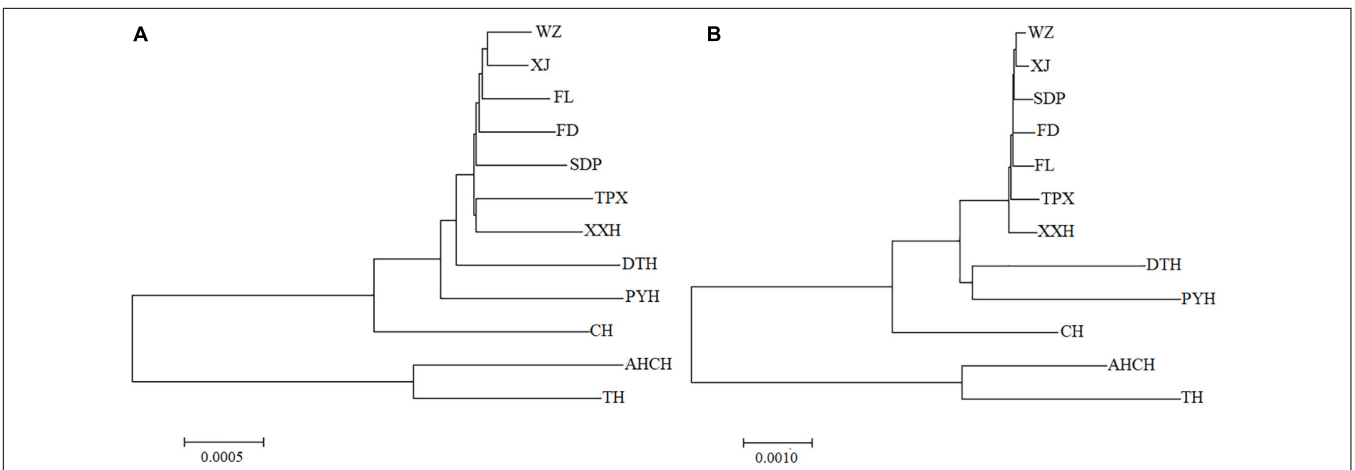


FIGURE 2 | Neighbor-joining tree of 12 *C. brachygnathus* populations based on K-2-P genetic distance of Cyt *b* **(A)** and D-Loop **(B)** genes.

the lower Yangtze River had similar genetic cluster components; PYH, DTH, and CH populations in the middle Yangtze River had similar genetic cluster components; SDP, TPX, XXH, XJ, WZ, FD, and FL populations in the Three Gorges Reservoir had similar genetic cluster components, and the Three Gorges Reservoir populations shared some genetic cluster components with middle Yangtze River populations. The PCA analysis showed that the lower Yangtze River populations were completely separated from

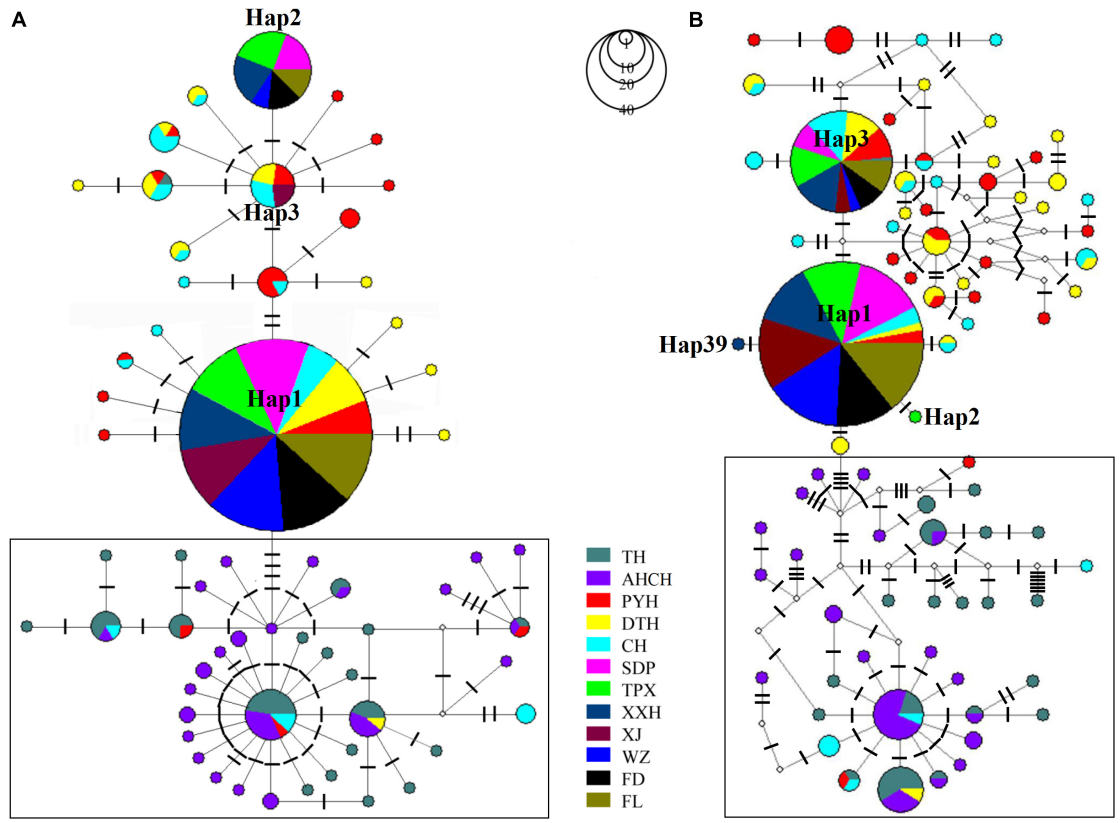


FIGURE 3 | Median-joining network of haplotypes of *C. brachygnathus* from 12 sites in the Yangtze River basin, China, inferred from Cyt b **(A)** and D-Loop **(B)** gene sequences. Each circle represents a haplotype, and its size is proportional to its frequency.

TABLE 4 | Gene flow between different groups based on mtDNA and SSR analyses.

Molecular marker	Cyt *b*	D-Loop	SSR
Three Gorges Reservoir—middle Yangtze River	3.20	1.00	1.55
Middle Yangtze River—lower Yangtze River	0.19	0.19	0.70
Three Gorges Reservoir—lower Yangtze River	0.11	0.08	0.38

the middle Yangtze River populations and the Three Gorges Reservoir populations; the middle Yangtze River populations and the Three Gorges Reservoir populations could also be clearly distinguished, but there were some overlapping areas (**Figure 5**). The gene flow between the Three Gorges Reservoir and the middle Yangtze River was significantly ($P < 0.05$) greater than those between the Three Gorges Reservoir and the lower Yangtze River (**Table 4**).

DISCUSSION

Population Genetic Structure and Invasion Sources

Identifying the invasion source and pathway of *C. brachygnathus* in the Three Gorges Reservoir was the top priority in this study, because the acquisition of this information is important for the prevention and control of *C. brachygnathus* in the Three Gorges Reservoir (Ficetola et al., 2008). Through this study, we speculated that the main invasion source of *C. brachygnathus* in the Three Gorges Reservoir was the middle Yangtze River instead of the lower Yangtze River. There were several results supporting this speculation. Firstly, the genetic differentiation, genetic distance between the Three Gorges Reservoir and the middle Yangtze River were smaller than between the Three Gorges Reservoir and the lower Yangtze River, and the gene flow result was opposite. Secondly, network of haplotypes indicated that there were shared haplotypes between the Three Gorges Reservoir populations and the middle Yangtze River populations, but no shared haplotypes between the Three Gorges Reservoir populations and the lower Yangtze River populations; the STRUCTURE analysis showed that the Three Gorges Reservoir populations only have some of the same genetic cluster components as the middle Yangtze River populations; PCA showed that the Three Gorges Reservoir populations only have some overlapping areas with the middle Yangtze River populations. Yang (2019) also found that the genetic distance of *C. brachygnathus* between the Three Gorges Reservoir populations and Dongtinghu Lake population was small. The reason for the middle Yangtze River population being the main invasion source, instead of the lower Yangtze River population, of *C. brachygnathus* in the Three Gorges Reservoir may be related to

TABLE 5 | Pairwise F_{ST} values based on SSR analysis of 384 samples of *C. brachygnathus* from 12 different sites in the Yangtze River Basin, China.

	TH	AHCH	PYH	DTH	CH	SDP	TPX	XXH	XJ	WZ	FD	FL
TH												
AHCH	0.02971*											
PYH	0.28764*	0.23085*										
DTH	0.29727*	0.24228*	0.00149									
CH	0.30233*	0.24850*	−0.00046	0.0054								
SDP	0.40955*	0.37237*	0.14514*	0.13291*	0.12093*							
TPX	0.40974*	0.37738*	0.16550*	0.14394*	0.15817*	0.05318*						
XXH	0.38759*	0.35168*	0.14686*	0.13114*	0.14705*	0.04877*	0.02026					
XJ	0.41927*	0.38665*	0.15975*	0.13951*	0.14720*	0.03663*	0.01826	0.01945				
WZ	0.39245*	0.35704*	0.13772*	0.11563*	0.13364*	0.03942*	0.00212	0.01036	0.00287			
FD	0.38955*	0.35231*	0.12256*	0.09911*	0.11353*	0.03418*	0.01832	0.00703	0.00969	−0.00222		
FL	0.39014*	0.35677*	0.15903*	0.13743*	0.15255*	0.05050*	0.01261	−0.00044	0.02319	0.00816	0.0053	

**P < 0.05 after Bonferroni correction.*

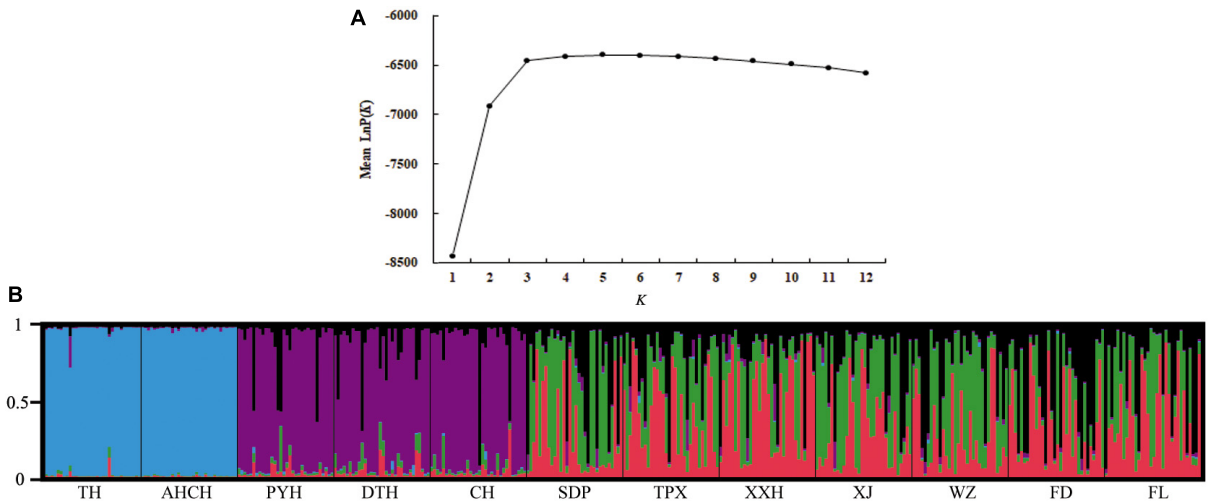


FIGURE 4 | STRUCTURE analysis based on genotypes at eight microsatellite loci from 384 unique *C. brachygnathus* individuals sampled from 12 different sites in the Yangtze River Basin, China. **(A)** Estimation of mean log probability LnP(K) for K = 1–12 using Structure Harvester. K = 5 is shown as the optimal value. **(B)** Proportions of each individual assigned to each inferred cluster. The five colors indicate five inferred distinct genetic clusters. Sampling sites of individuals are shown at the bottom.

geographic distance and the dispersal ability of *C. brachygnathus* (Radinger and Wolter, 2014; Radinger et al., 2017). Due to limited diffusion capacity, *C. brachygnathus* in the middle reaches of the Yangtze River can spread to the Three Gorges Reservoir, while *C. brachygnathus* in the lower Yangtze River hardly spread to the Three Gorges Reservoir.

As for the invasion pathway, we speculated that the most probable pathway for *C. brachygnathus* to have invaded the Three Gorges Reservoir was individuals from Poyang Lake, Dongting Lake, Changhu Lake, and other middle reaches of the Yangtze River actively moving upstream to the Three Gorges Reservoir through the operation of ship locks. When the ship was transported from under the dam to above the dam by cascade ship lock, the fish were also transported up. Lin et al. (2013) combined fish catch surveys and hydroacoustic detection to explore the fish passing ability of the ship locks of the Gezhouba and Three Gorges Dam; in the fish catch surveys of the two locks, a total of 21 species were identified, and the fish catch at the Gezhouba ship lock contained *C. brachygnathus*. Consistent with this, Xiang et al. (2015) used sonar (DIDSON) to detect and record the number of fish crossing up and down the Gezhouba ship lock and found that fish were able to pass the ship lock under natural conditions. Therefore, it is reasonable that *C. brachygnathus* spread to the Three Gorges Reservoir through the ship locks.

Genetic Diversity and Invasion Success

For many invasive species, the level of genetic diversity of an invaded area is often lower than that in the area of origin, because the invaded population is often produced from a few individuals from the original area; this is known as the founder effect (Allendorf and Lundquist, 2003; Crawford and Whitney, 2010; Peischl and Excoffier, 2015). In the present study, whether considering mitochondrial genes or microsatellites, the genetic diversity of *C. brachygnathus* in the non-invaded habitats (TH,

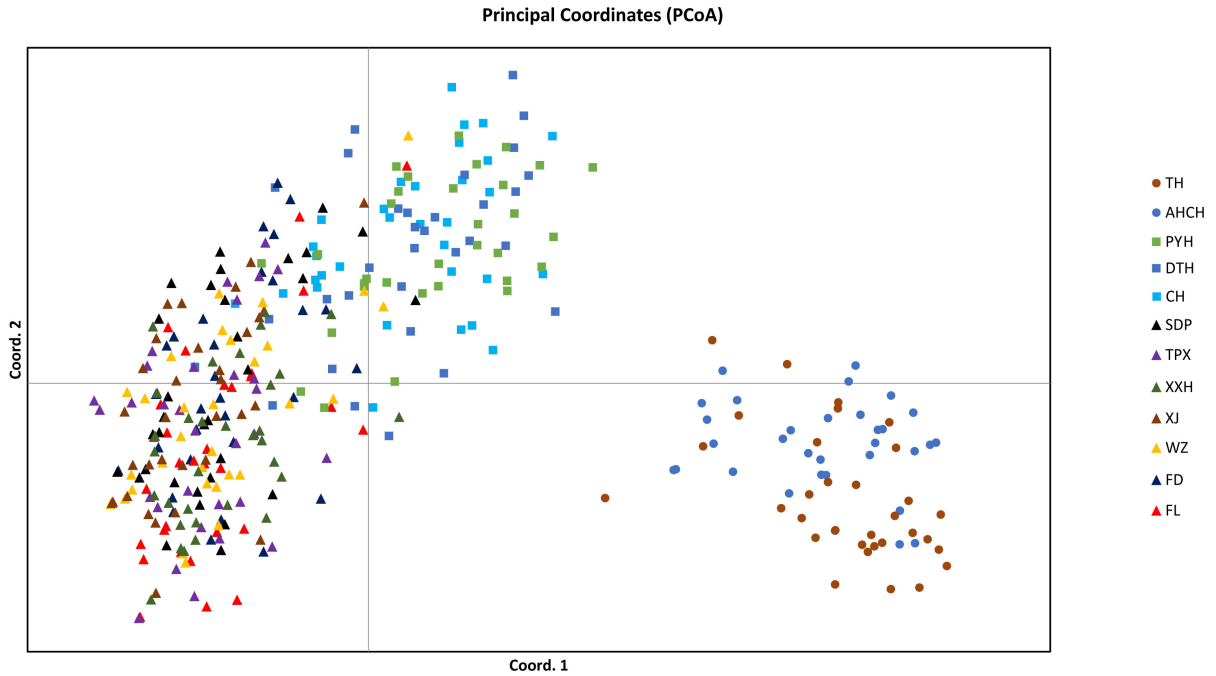


FIGURE 5 | The PCA showing the relationship of the PC1 and PC2 components of 12 *C. brachygnathus* populations in the Yangtze River Basin. Each triangle (square or circle) represents one sample corresponding to a specific population. The triangles represent the samples from the Three Gorges Reservoir; the squares represent the samples from the middle Yangtze River, and the circles represent the samples from the lower Yangtze River.

AHCH, PYH, DTH, CH) was significantly greater than that in the invaded habitats (SDP, TPX, XXH, XJ, WZ, FD, FL), indicating that the invaded populations experienced a founder effect. However, some studies have found that the genetic diversity of invasive species in the invaded region was not lower than that in the original area or was even higher than that in the area of origin (Frankham, 2005; Stepien et al., 2005). For example, Diez-del-Molino et al. (2013) found that Mosquitofish populations from the invaded Spanish streams had similar levels of genetic diversity as in the native American samples, suggesting these invasive populations did not appear to have undergone substantial loss of genetic diversity during the invasion process. The occurrence of this anomaly might be attributed to a single large number of individual invasions or to multiple introductions (Roman and Darling, 2007; Dlugosch and Parker, 2008). A single invasion of a large number of individuals makes the invading individuals include almost all the genetic diversity of the original population, and multiple introductions comprise genetic diversity of multiple ranges. Hybridization between individuals from different sources will also increase the genetic diversity of the population (Ellstrand and Schierenbeck, 2006).

The above discussion suggests that among the successful invasion cases, there are two contrasting patterns of genetic diversity comparisons of native and invaded ranges for different invasion cases. Therefore, what is the relationship between genetic diversity and successful invasion? It is generally believed that a high level of genetic diversity is beneficial to the settlement and maintenance of alien species in the new environment (Kolbe et al., 2004; Crawford and Whitney, 2010; Kanarek and Webb, 2010). Low levels of genetic diversity may be accompanied by reduced heterozygosity and inbreeding depression, restricting the population's evolutionary potential and ability to adapt to environmental changes and increasing the risk of population extinction (Altizer et al., 2003; Pinsky and Palumbi, 2014). Reed and Frankham (2003) conducted a meta-analysis of 34 research studies involving genetic diversity and population fitness and found that the level of population genetic diversity was significantly positively correlated with population fitness. However, in the present study, the genetic diversity of the invasive *C. brachygnathus* was relatively low due to founder effect. Why then has the *C. brachygnathus* invasion been so successful? We speculate that the successful invasion of *C. brachygnathus* could be related to its life history characteristics and its niche in the fish community of the Three Gorges Reservoir. *C. brachygnathus* has sufficient food, high fecundity, and strong ability to resist predation in the Three Gorges Reservoir, as introduced in the introduction, and these factors provide the species a competitive advantage over other fish. Therefore, the relationship between fish genetic diversity and successful invasion is not a simple causal relationship. The success of fish invasion is related not only to the genetic diversity but also to the life history characteristics of the invasive species and the environmental conditions.

CONCLUSION AND PROSPECT

This study suggests that the main invasion source of *C. brachygnathus* in the Three Gorges Reservoir was PYH,

DTH, CH, and other populations in the middle reaches of the Yangtze River. The invasion pathway may have been via moving upstream through the operation of ship locks. After *C. brachygnathus* invaded the Three Gorges Reservoir, a founder effect occurred, and the genetic diversity was significantly lower than in the native populations, but the number of *C. brachygnathus* in the Three Gorges Reservoir area had risen sharply.

Studies have shown that when alien species are introduced into a new area, if there is a significant difference between the new environment and the original environment, then the alien species may undergo rapid evolutionary changes under the pressure of natural selection caused by the new environment (Whitney and Gabler, 2008; Colautti and Lau, 2015). In addition, recent studies have shown that evolutionary events can occur on a relatively short time scale (Thompson, 1998; Whitney and Gabler, 2008). Therefore, we predict that *C. brachygnathus* in the Three Gorges Reservoir may evolve rapidly, and the genetic structure of each generation may change. Indeed, this study found that *C. brachygnathus* in the Three Gorges Reservoir contained some novel haplotypes and genotypes. In order to monitor the genetic dynamics of *C. brachygnathus* in the Three Gorges Reservoir with time and provide basic data for the prevention of *C. brachygnathus* in the Three Gorges Reservoir, we plan to conduct sampling and genetic testing of *C. brachygnathus* in the Three Gorges Reservoir every year. The inter-annual changes of the genetic structure of *C. brachygnathus* in the Three Gorges Reservoir can thus be examined on a longer time scale. In addition, more sensitive markers than mitochondrial gene and microsatellite markers, such as SNP, can be used for related research in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI (accession: OK172180–OK172317).

ETHICS STATEMENT

The animal study was reviewed and approved by the Jianghan University.

AUTHOR CONTRIBUTIONS

DZ contributed to sampling, performed the laboratory work, and wrote the manuscript. FX and WJ designed the study and revised the manuscript. BL and HL helped to data analysis. CL helped to laboratory work. XD and DC helped to revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (2018YFD0900903), the Research Project of China Three Gorges Corporation (2020049-ZHX), the National Natural Science Foundation of China (51779105 and 51979123), the Innovative Research Team Foundation of the Department of Education of Hubei Province, China (T2020034), and the Scientific Research Project of Jianghan University (2021kjzx006).

ACKNOWLEDGMENTS

We are grateful to Jin Cai and Jixin Yu for their help in *C. brachygnathus* sampling. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.783215/full#supplementary-material

REFERENCES

Allendorf, F. W., and Lundquist, L. L. (2003). Introduction: population biology, evolution, and control of invasive species. *Conserv. Biol.* 17, 24–30. doi: 10.1046/j.1523-1739.2003.02365.x

Altizer, S., Harvell, D., and Friedle, E. (2003). Rapid evolutionary dynamics and disease threats to biodiversity. *Trends Ecol. Evol.* 18, 589–596. doi: 10.1016/j.tree.2003.08.013

Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Bariche, M., Kleitou, P., Kalogirou, S., and Bernardi, G. (2017). Genetics reveal the identity and origin of the lionfish invasion in the Mediterranean Sea. *Sci. Rep.* 7:6782. doi: 10.1038/s41598-017-07326-1

Chen, Y. Y. (1998). *Fauna Sinica, Osteichthyes Cypriniformes II*. Beijing: Science Press.

Colautti, R. I., and Lau, J. A. (2015). Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Mol. Ecol.* 24, 1999–2017. doi: 10.1111/mec.13162

Crawford, K. M., and Whitney, K. D. (2010). Population genetic diversity influences colonization success. *Mol. Ecol.* 19, 1253–1263. doi: 10.1111/j.1365-294X.2010.04550.x

Diez-del-Molino, D., Carmona-Catot, G., Araguas, R. M., Vidal, O., Sanz, N., Garcia-Berthou, E., et al. (2013). Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). *PLoS One* 8:e82501. doi: 10.1371/journal.pone.0082501

Ding, R. H. (1994). *The Fishes of Sichuan, China*. Chengdu: Sichuan Publishing House of Science and Technology.

Dlugosch, K. M., and Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* 17, 431–449. doi: 10.1111/j.1365-294X.2007.03538.x

Du, F. K., Tang, Y. K., Yu, J. H., Su, S. Y., Yu, F., Li, J. L., et al. (2019). Development and validation of a transcriptome-based simple sequence repeats markers in *Coilia nasus*. *Pak. J. Zool.* 51, 1335–1341. doi: 10.17582/journal.pjz/2019.51.4.1335.1341

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Leveque, C., et al. (2006). Freshwater biodiversity: importance, threats,

status and conservation challenges. *Biol. Rev.* 81, 163–182. doi: 10.1017/S1464793105006950

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Ellstrand, N. C., and Schierenbeck, K. A. (2006). Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica* 148, 35–46. doi: 10.1073/pnas.97.13.7043

Estoup, A., and Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Mol. Ecol.* 19, 4113–4130. doi: 10.1111/j.1365-294X.2010.04773.x

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform.* 1, 47–50. doi: 10.1177/117693430500100003

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi: 10.1093/genetics/164.4.1567

Ficetola, G. F., Bonin, A., and Miaud, C. (2008). Population genetics reveals origin and number of founders in a biological invasion. *Mol. Ecol.* 17, 773–782. doi: 10.1111/j.1365-294X.2007.03622.x

Fish research laboratory of Institute of Hydrobiology of Hubei province. (1976). *The Yangtze River Fish*. Beijing: Science Press.

Frankham, R. (2005). Resolving the genetic paradox in invasive species. *Heredity* 94, 385–385. doi: 10.1038/sj.hdy.6800634

Gao, X., Fujiwara, M., Winemiller, K. O., Lin, P. C., Li, M. Z., and Liu, H. Z. (2019). Regime shift in fish assemblage structure in the Yangtze River following construction of the Three Gorges Dam. *Sci. Rep.* 9:4212. doi: 10.1038/s41598-019-38993-x

Gao, X., Zeng, Y., Wang, J. W., and Liu, H. Z. (2010). Immediate impacts of the second impoundment on fish communities in the Three Gorges Reservoir. *Environ. Biol. Fish.* 2010, 163–173. doi: 10.1007/s10641-009-9577-1

Goudet, J. (2001). *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices*. Available online at: http://www.unil.ch/izea/softwares/fstat.html (accessed Auguest 23, 2005).

Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361–372. doi: 10.2307/2532296

Haubrock, P. J., Pilotto, F., Innocenti, G., Cianfanelli, S., and Haase, P. (2021). Two centuries for an almost complete community turnover from native to non-native species in a riverine ecosystem. *Glob. Change Biol.* 27, 606–623. doi: 10.1111/gcb.15442

Huang, Z. L., and Wu, B. F. (2018). *Three Gorges Dam*. Berlin: Springer.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

Kanarek, A. R., and Webb, C. T. (2010). Allee effects, adaptive evolution, and invasion success. *Evol. Appl.* 3, 122–135. doi: 10.1111/j.1752-4571.2009.00112.x

Kolbe, J. J., Glor, R. E., Schettino, L. R. G., Lara, A. C., Larson, A., and Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431, 177–181. doi: 10.1038/nature02807

Kreyenberg, W., and Pappenheim, P. (1908). Ein beitrag zur kenntnis der fische der jangtze und seiner zuflüsse. *Sitzungsber. Ges. Naturf. Freunde Berlin* 1908, 95–109.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Levine, J. M. (2008). Biological invasions. *Curr. Biol.* 18, R57–R60. doi: 10.1016/j.cub.2007.11.030

Li, X. H., Huang, J., Filker, S., Stoeck, T., Bi, Y. H., Yu, Y. H., et al. (2019). Spatio-temporal patterns of zooplankton in a main-stem dam affected tributary: a case study in the Xiangxi River of the Three Gorges Reservoir, China. *Sci China Life Sci*. 62, 1058–1069. doi: 10.1007/s11427-018-9523-0

Liao, C. S., Chen, S. B., De Silva, S. S., Correa, S. B., Yuan, J., Zhang, T. L., et al. (2018). Spatial changes of fish assemblages in relation to filling stages of the Three Gorges Reservoir, China. *J. Appl. Ichthyol.* 34, 1293–1303. doi: 10.1111/jai.13798

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Lin, P. C., Brosse, S., Gao, X., Liu, C. C., and Liu, H. Z. (2013). Species composition and temporal pattern of fish passing through the navigation locks in the middle reach of Yangtze River: implications for fish conservation. *J. Appl. Ichthyol.* 29, 1441–1444. doi: 10.1111/jai.12362

Liu, P. Z., Liu, M. H., Lei, G. C., Zeng, Q., Li, Y. Y., and Bridgewater, P. (2021). Conservation of the Yangtze River Basin, China. *Oryx* 55, 331–332. doi: 10.1017/S0030605321000065

Liu, Y. L. (2008). *Species Identification and Biological Characteristics of Coilia brachygnathus in the Poyang Lake*. Master's thesis. Nanchang: Nanchang University.

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., and Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* 10, 689–710.

Meyers, M. A., Lin, A. Y. M., Lin, Y. S., Olevsky, E. A., and Georgalis, S. (2008). The cutting edge: sharp biological materials. *JOM* 60, 19–24. doi: 10.1007/s11837-008-0027-x

Morim, T., Bigg, G. R., Madeira, P. M., Palma, J., Duvernell, D. D., Gisbert, E., et al. (2019). Invasion genetics of the mummichog (*Fundulus heteroclitus*): recent anthropogenic introduction in Iberia. *PeerJ* 7:e6155. doi: 10.7717/peerj.6155

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Resour.* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Peischl, S., and Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Mol. Ecol.* 24, 2084–2094. doi: 10.1111/mec.13154

Pinsky, M. L., and Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Mol. Ecol.* 23, 29–39. doi: 10.1111/mec.12509

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Radinger, J., Essl, F., Hölker, F., Horký, P., Slavík, O., and Wolter, C. (2017). The future distribution of river fish: the complex interplay of climate and land use changes, species dispersal and movement barriers. *Glob. Change Biol.* 23, 4970–4986. doi: 10.1111/gcb.13760

Radinger, J., and Wolter, C. (2014). Patterns and predictors of fish dispersal in rivers. *Fish Fish.* 15, 456–473. doi: 10.1111/faf.12028

Reed, D. H., and Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conserv. Biol.* 17, 230–237. doi: 10.1046/j.1523-1739.2003.01236.x

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* 43, 223–225. doi: 10.2307/2409177

Roman, J., and Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol. Evol.* 22, 454–464. doi: 10.1016/j.tree.2007.07.002

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Stepien, C. A., Brown, J. E., Neilson, M. E., and Tumeo, M. A. (2005). Genetic diversity of invasive species in the Great Lakes versus their Eurasian source populations: insights for risk analysis. *Risk Anal.* 25, 1043–1060. doi: 10.1111/j.1539-6924.2005.00655.x

Stone, R. (2008). Three Gorges Dam: into the unknown. *Science* 321, 628–632. doi: 10.1126/science.321.5889.628

Tang, W. Q., Hu, X. L., and Yang, J. Q. (2007). Species validities of *Coilia brachygnathus* and *C. nasus taihuensis* based on sequence variations of complete mtDNA control region. *Biodivers. Sci.* 15, 224–231. doi: 10.1360/biodiv.060263

Thompson, J. N. (1998). Rapid evolution as an ecological process. *Trends Ecol. Evol.* 13, 329–332. doi: 10.1016/S0169-5347(98)01378-0

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Resour.* 4, 535–538. doi: 10.1111/j.1471-8286.2004.00684.x

Wei, N., Zhang, Y., Wu, F., Shen, Z. W., Ru, H. J., and Ni, C. H. (2021). Current status and changes in fish assemblages in the Three Gorges Reservoir. *Resour. Environ. Yangtze Basin* 30, 1858–1869.

Wei, Q. W., Ke, F. E., Zhang, J. M., Zhuang, P., Luo, J. D., Zhou, R. Q., et al. (1997). Biology, fisheries, and conservation of sturgeons and paddlefish in China. *Environ. Biol. Fish.* 48, 241–255. doi: 10.1023/A:1007395612241

Whitehead, P. J. P., Nelson, G. J., and Wongratana, T. (1988). *FAO species catalogue. Vol. 7. Clupeoid Fishes of the World (Suborder Clupeoidei). An Annotated and Illustrated Catalogue of the Herrings, Sardines, Pilchards, Sprats, Shads, Anchovies and Wolf-Herrings. Part 2: Engraulididae*. Rome: FAO Fisheries Synopsis.

Whitney, K. D., and Gabler, C. A. (2008). Rapid evolution in introduced species,'invasive traits' and recipient communities: challenges for predicting invasive potential. *Divers. Distrib.* 14, 569–580. doi: 10.1111/j.1472-4642.2008.00473.x

Xiang, J. W., Wang, C. F., Liu, D. F., Zhou, J. F., Feng, X., and Zhao, P. (2015). Temporal characteristics of fish passing through the Gezhouba ship lock. *Resour. Environ. Yangtze Basin* 24, 455–463.

Xiao, W. H., Zhang, Y. P., and Liu, H. Z. (2001). Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in east Asia. *Mol. Phylogenet. Evol.* 18, 163–173. doi: 10.1006/mpev.2000.0879

Xu, K. H., and Milliman, J. D. (2009). Seasonal variations of sediment discharge from the Yangtze River before and after impoundment of the Three Gorges Dam. *Geomorphology* 104, 276–283. doi: 10.1016/j.geomorph.2008.09.004

Yang, F. (2019). *Genetic Diversity Comparison of Coilia Brachygnathus and Neosalanx Taihuensis Between Populations in the Three Gorges Reservoir and Dongting Lake*. Master Dissertation. El Paso, TX: Southwest University.

Yu, A. Q., Shi, Y. H., and Deng, P. P. (2019). Microsatellite analysis of genetic diversity of wild population and selection of *Coilia nasus* in the Yangtze River. *Aquat. Sci. Technol. Inform.* 46, 121–125.

Zhang, H., Wu, G. G., Xie, P., Xu, J., and Zhou, Q. (2013). Role of body size and temporal hydrology in the dietary shifts of shortjaw tapertail anchovy *Coilia brachygnathus* (Actinopterygii, Engraulidae) in a large floodplain lake. *Hydrobiologia* 703, 247–256. doi: 10.1007/s10750-012-1370-z

Conflict of Interest: BL and WJ were employed by Three Gorges Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fgene.2022.795850

Whole-Genome Resequencing to Evaluate Life History Variation in Anadromous Migration of *Oncorhynchus mykiss*

Erin E. Collins[1]*, *Nicolas Romero*[2], *Joseph S. Zendt*[2] and *Shawn R. Narum*[1]

[1]Hagerman Genetics Laboratory, Columbia River Inter-Tribal Fish Commission, Hagerman, ID, United States, [2]Yakama Nation Fisheries, Yakima/Klickitat Fisheries Project, Klickitat, WA, United States

OPEN ACCESS

Edited by:
Narongrit Muangmai,
Kasetsart University, Thailand

Reviewed by:
Timothy D. Leeds,
United States Department of Agriculture (USDA), United States
Matt Hale,
Texas Christian University, United States

***Correspondence:**
Erin E. Collins
ecollins@critfc.org

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 15 October 2021
Accepted: 24 January 2022
Published: 15 March 2022

Citation:
Collins EE, Romero N, Zendt JS and Narum SR (2022) Whole-Genome Resequencing to Evaluate Life History Variation in Anadromous Migration of Oncorhynchus mykiss. Front. Genet. 13:795850. doi: 10.3389/fgene.2022.795850

Anadromous fish experience physiological modifications necessary to migrate between vastly different freshwater and marine environments, but some species such as *Oncorhynchus mykiss* demonstrate variation in life history strategies with some individuals remaining exclusively resident in freshwater, whereas others undergo anadromous migration. Because there is limited understanding of genes involved in this life history variation across populations of this species, we evaluated the genomic difference between known anadromous (n = 39) and resident (n = 78) *Oncorhynchus mykiss* collected from the Klickitat River, WA, USA, with whole-genome resequencing methods. Sequencing of these collections yielded 5.64 million single-nucleotide polymorphisms that were tested for significant differences between resident and anadromous groups along with previously identified candidate gene regions. Although a few regions of the genome were marginally significant, there was one region on chromosome Omy12 that provided the most consistent signal of association with anadromy near two annotated genes in the reference assembly: COP9 signalosome complex subunit 6 (CSN6) and NACHT, LRR, and PYD domain–containing protein 3 (NLRP3). Previously identified candidate genes for anadromy within the inversion region of chromosome Omy05 in coastal steelhead and rainbow trout were not informative for this population as shown in previous studies. Results indicate that the significant region on chromosome Omy12 may represent a minor effect gene for male anadromy and suggests that this life history variation in *Oncorhynchus mykiss* is more strongly driven by other mechanisms related to environmental rearing such as epigenetic modification, gene expression, and phenotypic plasticity. Further studies into regulatory mechanisms of this trait are needed to understand drivers of anadromy in populations of this protected species.

Keywords: *Oncorhynchus*, steelhead, anadromy, whole-genome resequencing, Klickitat River

1 INTRODUCTION

Variation in life history traits is often advantageous to the persistence of natural species, and salmonids retain diverse life history types to contend with environmental stochasticity (e.g., Moore et al., 2014). Species such as *Oncorhynchus mykiss* demonstrate variation in migratory life history with some individuals remaining exclusively resident in freshwater, whereas others exhibit anadromous migration from freshwater rearing environments to the ocean before returning to natal areas to spawn. Populations can consist of primarily anadromous or resident fish, or there can be a mixture of resident and anadromous individuals within a single population. Anadromy is rare, <1% of all fish species are anadromous (McDowall 1988), because the biological cost includes challenges of surviving in both fresh and saltwater habitats (e.g., physiological changes to regulate ions, long distance migrations, and predation rates) and must be outweighed by benefits such as improved productivity and access to seasonally available resources to feed and spawn. Because of smoltification and migration of anadromous fish, reproduction is delayed for adults; however, they benefit from large size and high fecundity gained from highly productive marine feeding habitat. Ultimately, life history development is a complex trait and is influenced by genetics, environment, and individual fitness (Fleming and Reynolds 2004). Variation in the degree of anadromy or residency of *O. mykiss* populations impacts the abundance, diversity, resilience, structure, and productivity of *O. mykiss* populations (Waples et al., 2008). There are greater benefits for female *O. mykiss* to migrate and grow to a larger size before spawning compared to males. Larger female *O. mykiss* experience increased fecundity, increased egg size, improved redd site protection, can dig deeper into the substrate, and can attract more mates (Quinn et al., 2011). Whereas, male *O. mykiss* are more likely than females to be resident due to an option for males to spawn at a small size with the "sneaky/sneaker" male tactic (Fleming and Reynolds 2004), and this variation in males was the focus of this study.

Oncorhynchus mykiss are of immense economic, ecological, and cultural importance throughout their range. Populations of *O. mykiss* are considered partially migratory due to the presence of both anadromous, "steelhead," and resident, "rainbow trout," forms of the species. These life history types of *O. mykiss* are sympatric and can produce offspring of either life history type (Christie et al., 2011; Courter et al., 2013; Sloat and Reeves 2014). Anadromy of *O. mykiss* are initiated by a combination of environmental and genetic cues and the genetic component is currently not well understood across much of the geographic range of this species. A continued presence of both resident and anadromous *O. mykiss* is critical to population resilience to environmental or anthropogenic disturbances (Moore et al., 2014).

Previous studies have identified candidate genes or regions in the genome associated with the trait for anadromy in steelhead, but association has been limited to particular populations. For example, Pearse et al. (2019) and Martínez et al. (2011) detected a "migration supergene" in an inversion on chromosome Omy05, where the ancestral form is associated with anadromous steelhead and the derived form is associated with resident rainbow trout. The signal from the chromosome Omy05 inversion was significant in coastal populations from southern portions of the range but had low variability for other populations of steelhead (inland or northern latitudes). Studies of anadromy in specific populations of *O. mykiss* have identified putative candidate genes associated with resident versus migratory life history (e.g., Nichols et al., 2008; Narum et al., 2011; Hale et al., 2013), but these candidates have not been directly validated in further studies. In addition, previous studies (Nichols et al., 2008; Narum et al., 2011; Hale et al., 2013; Hecht et al., 2013) detected outlier loci of small effect throughout the genome for anadromy, indicating an extremely polygenic trait, and very few of the same loci were detected in multiple analyses and little overlap was observed between populations tested, indicating a lack of a conserved gene region or high environmental plasticity for the trait. Thus, studies to date indicate that the genetic basis of anadromy is controlled in a population specific manner. However, there is currently a paucity of data from many populations, limiting our understanding of the relative importance of shared alleles and population specific anadromy.

Steelhead in the Klickitat River, WA, USA, are geographically and genetically intermediate between coastal and inland genetic lineages (Blankenship et al., 2011; Collins et al., 2020). The geography along the Klickitat River and its tributaries is diverse and includes waterfalls that act as partial barriers to migration, leading to reproductive and genetic isolation for populations in some areas (Narum et al., 2006; Narum et al., 2008). Another genetic influence on Klickitat River *O. mykiss* populations has been the stocking of summer-run hatchery steelhead, known as Skamania steelhead, since 1961, and the Skamania steelhead may have overlapping spawning timing with natural origin summer steelhead, but with limited gene flow (Narum et al., 2006). Skamania steelhead were originally derived from steelhead from the Washougal and Klickitat Rivers and have been strongly selected for early spawning times (Crawford 1979). Resident *O. mykiss* occur throughout substantial portions of this drainage, especially in headwater tributaries with lower flows and water temperature than large tributaries or the mainstem Klickitat River (Narum et al., 2008; Evans et al., 2021). Because Klickitat River anadromous steelhead are listed as threatened and are protected under the Endangered Species Act, identifying that the genetic component of anadromy may be a useful conservation tool to assist with recovery of steelhead in this system.

In this study, we aimed to use whole-genome resequencing to assess underlying genetic mechanisms for the trait of male anadromy in *O. mykiss* from the Klickitat River. This study combined high marker density throughout the genome and phenotypic data from long-term tracking of individuals to test for regions of the genome associated with resident versus anadromous migration patterns. Further, markers from previously identified candidate genes were tested for differences among collections. A genetic sex marker on the basis of the sexually dimorphic on the Y chromosome region (sdY; Yano et al., 2013) was used to identify males and females, enabling analyses to focus on sex specific migration patterns in

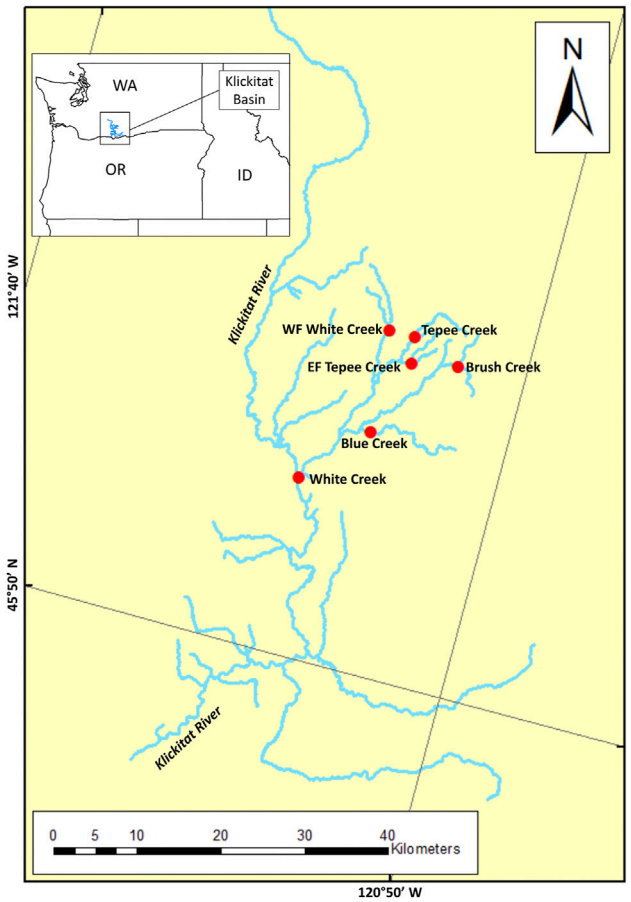


FIGURE 1 | Map of sample sites within the White Creek tributary of the Klickitat River, WA, USA, marked with a red circles. Individual samples within this tributary were identified as either resident (n = 78) or anadromous (n = 39) on the basis of tracking migratory fish patterns with passive integrated transponder (PIT) tags.

males to avoid sex bias when comparing anadromy life traits due to the biological constraint of rare resident female rainbow trout.

2 MATERIALS AND METHODS

2.1 Sample Collection and Phenotypes

Samples of juvenile *O. mykiss* were collected from the Klickitat River *via* electrofishing from the following tributaries of the White Creek drainage: White Creek, West Fork White Creek, Blue Creek, Brush Creek, Tepee Creek, and East Fork Tepee Creek (**Supplementary Table S1** and **Figure 1**). All individual *O. mykiss* were captured as juveniles (parr appearance) in multiple sites of the White Creek drainage between 2009 and 2019 and were inserted with a 12.5-mm full-duplex passive integrated transponder (FDX-B PIT) tag (Biomark, Boise, ID, USA) to track individual migration patterns and a non-lethal fin clip was collected for genetic analyses. Fish were tracked for multiples years to determine migratory classification as either resident or anadromous for genomic analyses.

Sampling was ongoing and conducted every year with marked fish continuously monitored *via* passive mark recapture methods. Captured fish were initially scanned with a handheld PIT tag scanner to identify previously PIT-tagged fish (recaptures) from non-tagged fish at each sample site. A PIT tag was inserted in each non-tagged *O. mykiss* ≥ 65 mm. Sample date, PIT tag code, and biometric data were recorded from each fish. In addition, a non-lethal fin clip for genetic analysis was collected from first time captured fish. Each sampled fish was returned to the stream at their capture location. Movement patterns were tracked through a combination of active and passive mark-recapture methods. Electrofishing surveys were conducted at each site annually to actively monitor fish. Electrofishing was the method used to both capture fish for marking and to manually recapture previously tagged fish. We used electrofishing as a method to classify resident fish by establishing a temporal record of recapture events for a given individual.

A network of PIT tag interrogation arrays in the Klickitat River and Columbia River were used to passively monitor migratory behavior. PIT tag interrogation arrays located in tributaries to the Klickitat River were designed to interrogate the entire wetted channel width *via* channel spanning PIT tag arrays. The mainstem Klickitat River floating array interrogated approximately two-thirds of the mainstem wetted channel. Sites varied from having a minimum of two arrays to a maximum of three arrays. Depending on stream width, arrays consisted of a single antenna to up to three adjacent antennas. Arrays were configured by placing antenna arrays perpendicular to the stream channel. PIT tag arrays were spaced five to 10 m apart along the longitudinal stream axis. For each fish detection, the PIT tag code, date/time stamp, and unique antenna number were recorded to a data logger buffer. Directionality of movement and temporal movement patterns were subsequently determined from these data. Juvenile bypass interrogation arrays at Bonneville Dam operate continuously, except for a short non-operational period in the winter. In addition, an array is towed through the Columbia River estuary to detect PIT-tagged fish throughout the year. All PIT tag interrogation sites were operational during the juvenile out-migration period.

Individual O.mykiss samples were classified as either resident or anadromous based on multiple years of mark-recapture sampling and monitoring. Anadromous fish were identified as those that successfully outmigrated from the Klickitat River and were either last detected in the mainstem Columbia River presumably headed to the ocean or returning from the ocean. Fish were classified as freshwater residents if they remained in freshwater for a minimum of 2 years between initial and final recapture events and did not register at any of the downstream PIT tag arrays.

2.2 Molecular Methods

Prior to whole-genome resequencing, specimens were genotyped with a panel of single-nucleotide polymorphism (SNP) markers to identify sex of samples and to account for kinship and population structure. A Chelex 100 method (Sigma-Aldrich, St. Louis, MO) was followed to extract DNA from the tissue of a total of 198 *O. mykiss* individuals (**Supplementary Table S1**). The genotyping-in-thousands by sequencing method (GT-seq; Campbell et al., 2015) was employed for all samples. Samples

were genotyped with a panel of 376 SNPs (Collins et al., 2020) that were a mix of putatively neutral and adaptive markers, a sex marker, and markers that identify closely related species of cutthroat trout (*O. clarkii*). Briefly, our study followed standard GT-seq methods that implemented two rounds of polymerase chain reaction (PCR) to first amplify targeted SNPs and then add dual barcodes for the identification of all individuals in downstream analyses. Following the barcoding step, the concentration of each sample was normalized and then pooled into a library. Once the library was prepared, it was sequenced on an Illumina NextSeq 500 instrument and then was genotyped with scripts from Campbell et al. (2015). As a quality control step, all samples and loci with ≥10% missing genotypes were removed from further analyses. Genotype data were retained when >90% loci successfully genotyped and had an estimated <0.5% genotyping error based on replicate genotyping of 10% of samples.

To select the final samples for the whole-genome resequencing step, samples were first analyzed and filtered according to biases from kinship, population structure, and sex imbalances among samples. The sex of individuals was determined with the *O. mykiss* sex marker (OmyY1_2SEXY; Brunelli et al., 2008) with the intent to split individuals into four groups (resident male, anadromous male, resident female, and anadromous female) for statistical analysis. This approach was intended to account for sex specific differences in migration behavior and isolate confounding signals of sex-linked genes including the known sdY region on chromosome Omy29 (Gao et al., 2021). Kinship and population biases were tested at neutral loci with a compressed linear model (Zhang et al., 2010) that used a kinship matrix produced with the VanRaden algorithm (VanRaden et al., 2008) along with three principal components and were implemented in Genome Association and Prediction Integrated Tool (GAPIT-R package; Lipka et al., 2012).

Our approach to genome resequencing was to target low read depth for individually barcoded samples and then to combine sequence reads from individuals of the same phenotype for further analyses of the four groups: resident male, anadromous male, resident female, and anadromous female. To prepare the libraries for whole-genome resequencing, individuals were barcoded, and a NEBNext Ultra enzymatic fragmentation protocol was followed (Horn et al., 2020). Briefly, we extracted DNA according to a Chelex extraction method (Sweet et al., 1996). Individual sample DNA concentrations were quantified with pico-green fluorescence on a Tecan M200 (Männedorf, Switzerland) to normalize the quantity of DNA from each sample within two standard deviations of the mean concentration. After barcoding individuals, the DNA samples were combined to avoid batch effects and normalized DNA was fragmented with NEBNext double-stranded DNA fragmentase, which fragments randomly throughout the genome. Fragments of 400–600 base pairs (bp) were size-selected, and then, ends of DNA strands were repaired with NEBNext end prep. NEBNext adaptors were ligated to the repaired and size-selected fragments and cleaned with SpriSelect Beads. We implemented PCR amplification and cleaned the PCR product with SpriSelect Beads. The quantity and quality of each library were assessed using quantitative PCR and a Tape Station with a DNA High Sensitivity D1000 kit and then normalized prior to sequencing. Final libraries were sequenced on an Illumina NextSeq 500 sequencer, targeting 500 million paired-end reads per library.

2.3 Statistical Analyses

Barcodes were useful to normalize sequence reads for individual samples, but samples were subsequently pooled for analyses by migratory traits to provide adequate sequencing depth to identify SNPs and estimate allele frequencies for each phenotypic group. The whole-genome resequencing data were analyzed with a pipeline, called PoolParty (Micheletti and Narum 2018), that can handle either pooled or individually barcoded sequence data. The PoolParty pipeline consists of bash scripts that utilize open-source packages to align to a reference genome and filter raw whole-genome resequencing data. In this study, sequence reads were aligned to *O. mykiss* reference assembly GCA_013265735.3 (Gao et al., 2021). Next, the pipeline assesses filtered alignments with BBMap (Bushnell 2016), bwa mem (Li 2013), samblaster (Faust and Hall 2014), Picard Tools (Broad Institute), and bcftools (Li et al., 2011). Mapped-read coverage statistics were determined to evaluate mapping efficiency to the reference genome with mean depth of coverage and proportion of the reference assembly genome covered by mapped reads and filtered accordingly. To filter the SNPs that were used in downstream analyses, we assessed coverage and minor allele frequency (MAF). When SNPs had coverage either below 15X or above 100X, or MAF below 0.05, they were filtered out of analyses. Pairwise analyses, such as the fixation index (F_{ST}), sliding window FST (sFST), Fisher's exact test (FET), and local score (Fariello et al., 2017), were conducted for resident versus anadromous groups with Popoolation2 scripts (Kofler et al., 2011). To determine significance of difference between SNPs in resident and anadromous groups, we used false discovery rate (FDR) corrected *p*-values from FET analyses of allele frequencies for individual SNPs. A QQ plot analysis was conducted to evaluate the potential bias of association tests and effects of covariates.

To test for polygenic signals of association with anadromy, we analyzed the data set with an R package called Random Forest (RF; Liaw and Wiener 2002). This RF package is a machine learning algorithm that provides a nonparametric framework to identify epistasis through identification of a group of loci predictive of phenotypic traits with significant differences in allele frequencies that better predict the trait than a single locus (Brieuc et al., 2018). To detect these loci, a "forest" of classification or regression trees is partitioned recursively by randomly sampling, with replacement, creating a training subset of data, and a random selection of a subset of predictors to find which partition of loci best predicts the trait by minimizing within-group variance or error. The RF algorithm first identifies the best predictive loci partition as the first node of the "tree" and then continues to randomly subset the remaining predictors. Selections of partitions were random to avoid bias and curtail differences within and between the "trees" (Goldstein et al., 2011; Rokach 2016). An error rate, a proportion of the variation explained (PVE), and an importance value were

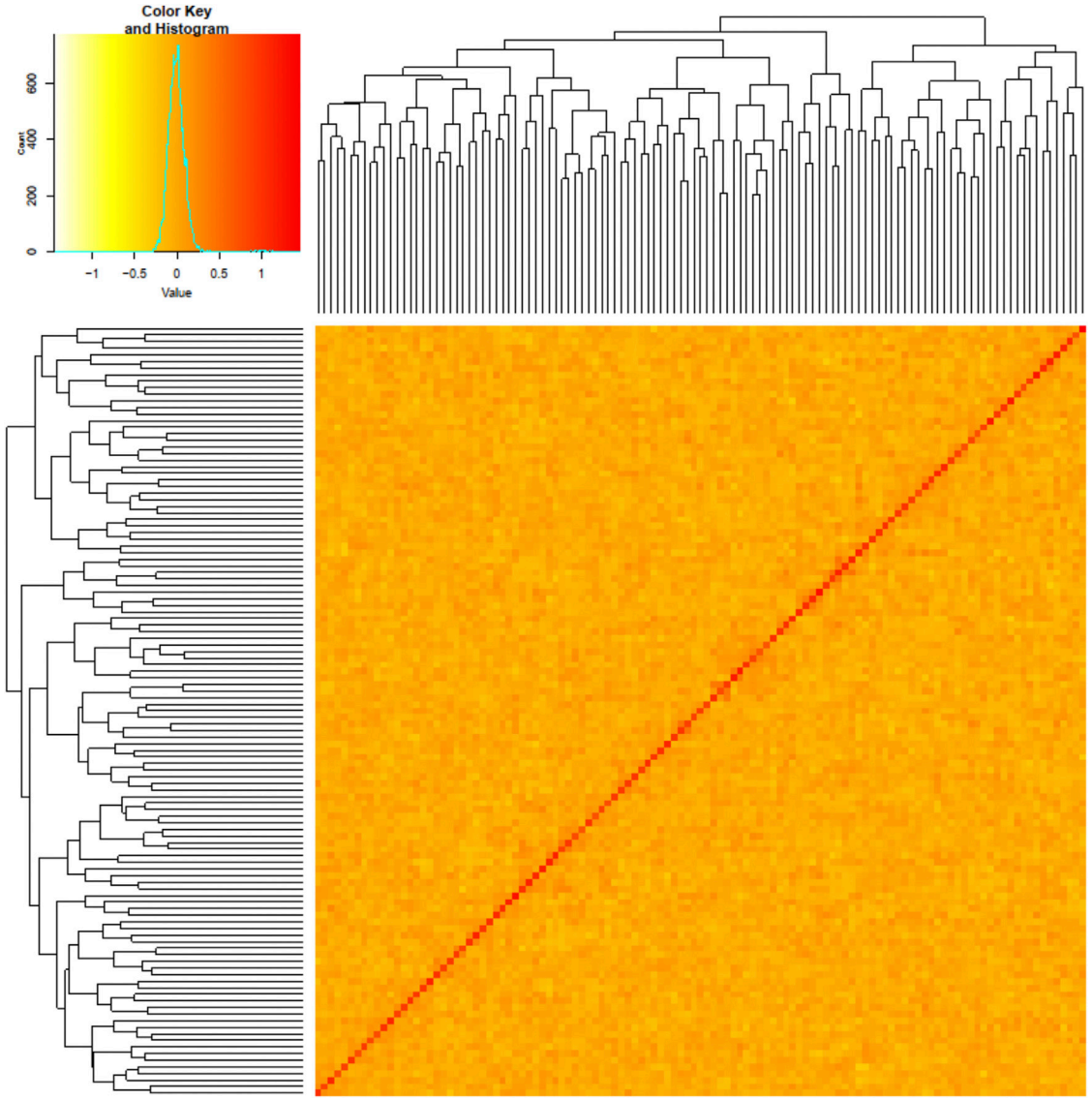


FIGURE 2 | Kinship matrix for 117 samples with darker colors and higher kinship values (e.g., kinship values ≥ 1 in this population) representing higher kinship between individuals. The diagonal represents relatedness values of each individual to itself as a reference for high kinship patterns. A phylogram of individual relatedness among samples is shown on both axes.

calculated with a subset of samples left out of the "tree" building steps as training data. The downside to random selections is that, if an important predictor is removed, then the predictive power of the "tree" will decrease; this can be belied with analyses of many "trees" and importance values of partitions across the "forest," instead of relying too heavily upon a single "tree".

To examine previously identified candidate genes for anadromy, we analyzed genotypes from candidate markers that were included in the GT-seq panel across phenotypic groups. This included several markers from the genome region described by Pearse et al. (2019) for large inversions on chromosome Omy05 of coastal California *O. mykiss* populations associated with the anadromy phenotype. In addition, we tested the three candidate loci from Narum et al. (2011) that had previously shown significant association with anadromy in previous analyses in *O. mykiss* from the Klickitat River. For these candidate markers that were genotyped with GT-seq, we assessed the MAF across all individuals and tested for differences in allele and genotype frequencies for anadromous and resident groups with FDR correction for multiple tests.

3 RESULTS

Results from genotyping the initial GT-seq panel were used to select the individuals for whole-genome resequencing. For the samples genotyped with GT-seq, the average genotype error rate was 0.51% and individuals missing ≥10% genotypes were

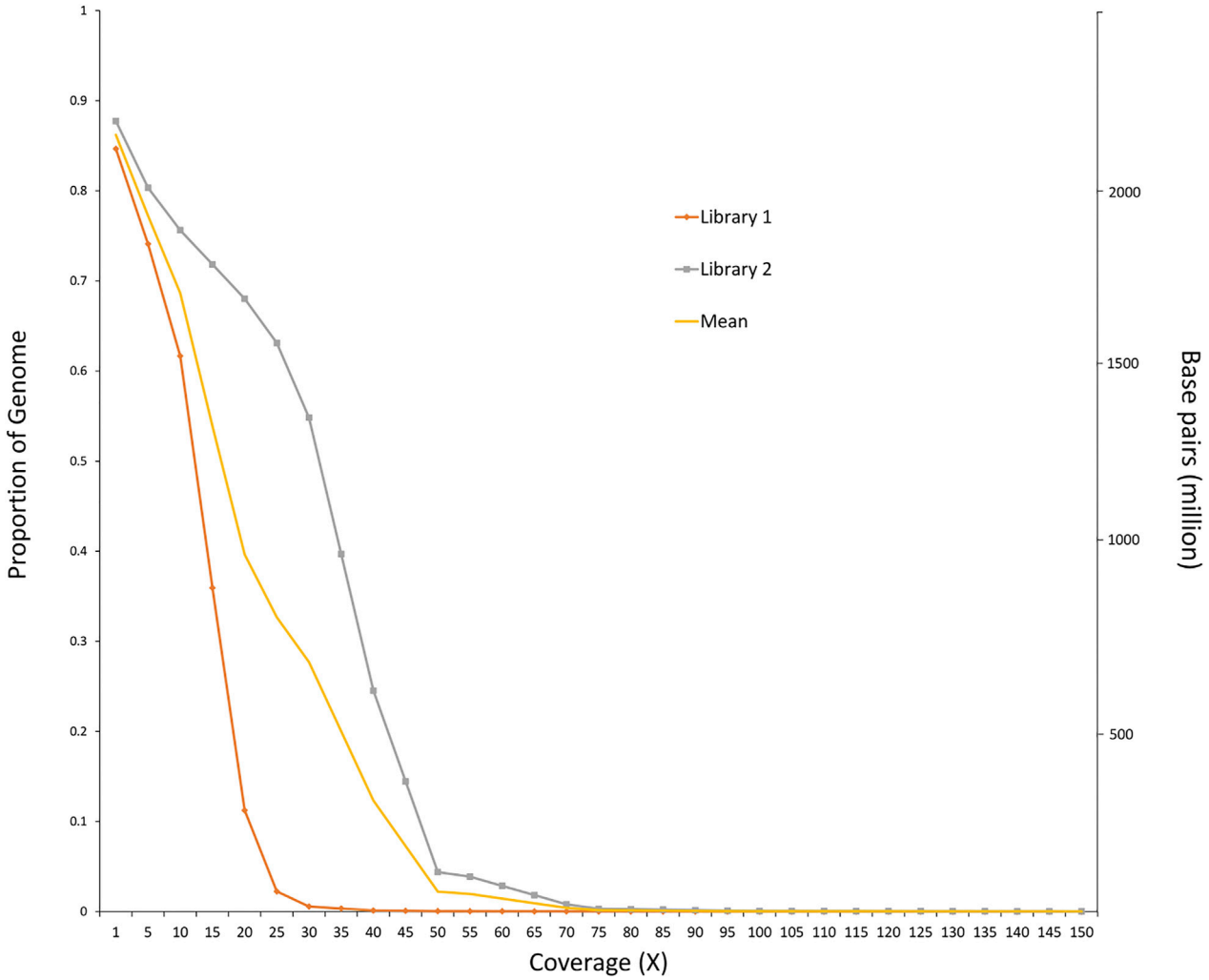


FIGURE 3 | Proportion of genome covered at different sequence read depths. The orange line represents Library 1, which consists of 39 anadromous males. The gray line represents library 2, which consists of 78 resident males.

excluded from analyses. Of the 198 individuals genotyped, three failed, leaving 195 for further analyses. Of those 195 individuals, the marker for genetic sex identified 117 males (78 resident and 39 anadromous) and 78 females (3 resident and 75 anadromous). This result reflected lower variation in migration life history for females than males (**Supplementary Table S1**) and a limited sample size of resident females that excluded intended analyses within the female groups (resident vs. anadromous). Analyses for potential kinship bias or population structure revealed minimal signal among the samples (**Figure 2**), and no individuals were excluded from further analyses.

The final sample set included 117 male *O. mykiss* that were split into two phenotypic groups with 39 anadromous (4 returning to spawn and 35 out-migrating) steelhead and 78 resident rainbow trout. Because there were insufficient resident females (n = 3) to test for migratory differences within females (resident vs. anadromous), an attempt was made to combine females with males in their respective phenotypic group even though this caused a severe imbalance in sex ratio between groups of residents (3 female, 78 male; 3.7% female) versus anadromous (75 female, 39 male; 67.8% female). Imbalanced sex ratios in analyses were expected to lead to detection of sex-linked regions as demonstrated in previous studies (Benestan et al., 2017) but were completed for thoroughness. The average sequence coverage for the anadromous males was 0.36× per individual and was 0.43× per individual for the resident males (**Figure 3**). After filtering, a total of 5.64 million SNPs were included for further tests for differences in allele frequency between phenotypic groups. Analyses with the PoolParty pipeline included results for F_{ST}, FET, sliding window F_{ST}, and local score analyses. Analyses of allele frequencies among the resident and anadromous groups indicated that six markers on multiple chromosomes were significant after Bonferroni correction with FET (**Figure 4A**) including those on chromosomes Omy06, Omy08, Omy12, Omy26, and Omy29 (**Table 1**). However,

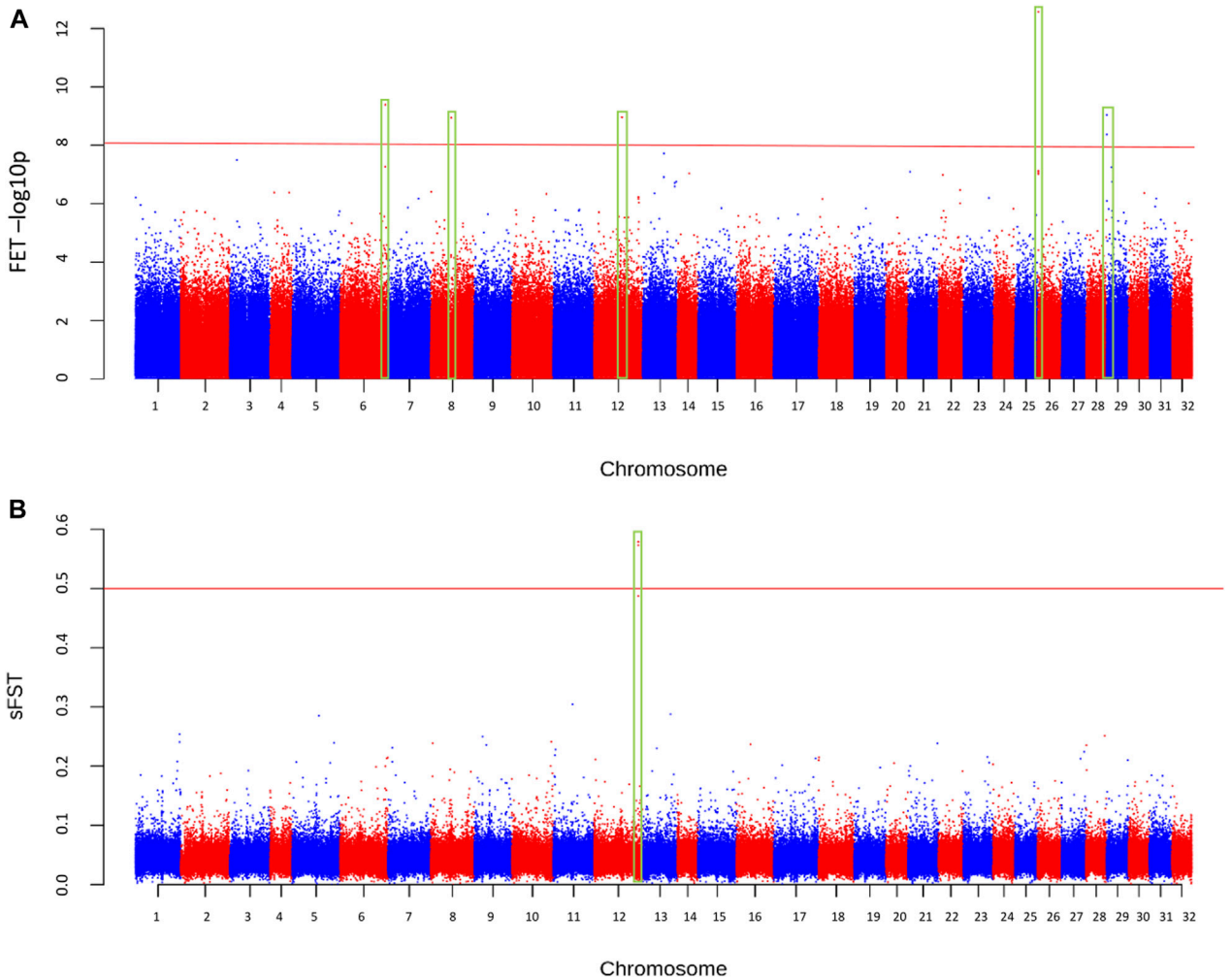


FIGURE 4 | Manhattan plot of differences in allele frequencies between resident and anadromous collections across the genome based on **(A)** significance from Fisher's exact test (FET), and **(B)** sliding window F_{ST} (sF_{ST}). The minimum coverage threshold was 15 reads and the minor allele frequency was > = 0.05.

only SNPs on chromosome Omy12 were significant among groups with sliding window F_{ST} and were considered the strongest candidate region for anadromy (**Figure 4B**). Significant SNPs on chromosome Omy12 were located 28.6–43.6 kb from the nearest gene, NACHT, LRR, and PYD domain–containing protein 3 (NLRP3), detected with the sliding window F_{ST} analysis (**Table 1**). The gene positioned closest (10.9 kb) to the significant SNP detected with FET analysis on chromosome Omy12 was the COP9 signalosome complex subunit 6 (CSN6) (**Table 1**). Analyses of combined sexes for resident vs. anadromous (with highly imbalanced sex ratios) yielded a strong signal near the sdY sex determining region on chromosome Omy29 as expected (**Supplementary Figure S5**).

Further analyses that account for physically linked SNPs (local score) or polygenic association (RF) provided no consistent signals of association and thus were considered false positive or unreliable results. Analyses with local score revealed many small but significant peaks on every chromosome of the genome (**Supplementary Figure S1**). Because local score is highly susceptible to false positives due to multiple testing effects of linked SNPs, a stringent critical value (p-value < 0.001 after Bonferroni correction) was applied and indicated that significant peaks occurred on several chromosomes (**Supplementary Figure S2**). We also searched for polygenetic association for epistatic loci with RF methods and significant loci were detected in multiple trees. However, candidates were not repeatable across replicate trees in the analyses and thus were discounted as reliable signals of polygenic association (**Supplementary Figure S4**). The QQ plot displayed most SNPs falling along the line representative of the null hypothesis or SNPs with no association to the anadromy trait (**Supplementary Figure S6**). The QQ plot also had loci that display deviation from the null hypothesis line, including every chromosome Omy12 marker identified with the other analyses, which were clustered even further from the null hypothesis line and above the stringent significance threshold of observed sF_{ST} > 0.5 (**Supplementary Figure S6**).

TABLE 1 | Genomic position and closest gene annotations for significant markers from Fisher's Exact Test (FET) and sliding window Fst (sFst) analyses on chromosome (Chr) Omy12.

SNP	Chr	Position	FET *p*-value value	sF_{ST} Differentiation value	Source	Closest Gene(s)	Possible Gene Function
1326266	6	95913987	4.03E-10	NA	PoolParty - fet	Leucine-rich repeat extensin-like protein 1	Leucine-rich repeat
1661566	8	43382768	1.14E-09	NA	PoolParty - fet	MYC proto-oncogene, bHLH transcription factor b	Leucine zipper, Myc, Myc-type, basic helix-loop-helix (bHLH) domain, transcription regulator Myc, N-terminal
2557802	12	58504189	1.10E-09	NA	PoolParty - fet	COP9 signalosome complex subunit 6	Posttranslational modification, protein turnover, chaperones, cell wall/membrane/envelope biogenesis
4778210	26	1643096	2.62E-13	NA	PoolParty - fet	Calcineurin-like phosphoesterase domain, ApaH type	Carbohydrate transport and metabolism, posttranslational modification, protein turnover, chaperones, mobilome, prophages, transposons
5119622	29	77226	8.99E-10	NA	PoolParty - fet	Anoctamin	Protein dimerization activity, integral component of membrane
5119623	29	77228	4.23E-09	NA	PoolParty - fet	Anoctamin	Protein dimerization activity, integral component of membrane
204161	12	93625000	NA	0.487	PoolParty - Sfst	NACHT, LRR, and PYD domain–containing protein 3	Leucine-rich repeat, transcription, signal transduction mechanisms, inorganic ion transport and metabolism
204162	12	93630000	NA	0.573	PoolParty - Sfst	NACHT, LRR, and PYD domain–containing protein 3	Leucine-rich repeat, transcription, signal transduction mechanisms, inorganic ion transport and metabolism
204163	12	93635000	NA	0.579	PoolParty - Sfst	NACHT, LRR, and PYD domain–containing protein 3	Leucine-rich repeat, transcription, signal transduction mechanisms, inorganic ion transport and metabolism
204164	12	93640000	NA	0.579	PoolParty - Sfst	NACHT, LRR, and PYD domain–containing protein 3	Leucine-rich repeat, transcription, signal transduction mechanisms, inorganic ion transport and metabolism

TABLE 2 | Chromosome (Chr) Omy05 markers from Pearse et al.(2019) and candidate genetic markers from Narum et al. (2011) were analyzed with minor allele frequencies (MAF) and Fisher's Exact Test (FET) for the samples for this study.

Genetic marker name	Chr	Total MAF	Anadromous MAF	Resident MAF	Homozygous MAF	Heterozygous MAF	Homozygous Major Allele Frequency	Inversion	FET *p*-value
OmyR19198Pearse	5	0.021	0.026	0.019	0	0.043	0.957	1	1
OmyR40252Pearse	5	0.021	0.026	0.019	0	0.043	0.957	1	1
OmyR14589Pearse	5	0.021	0.026	0.019	0	0.043	0.957	1	1
Omy_RAD23894-58	5	0.137	0.179	0.115	0.009	0.256	0.735	1	0.202
OmyR24370Pearse	5	0.021	0.026	0.019	0	0.043	0.957	1	1
Omy_bcAKala-380th	5	0.462	0.449	0.468	0.248	0.427	0.325	2	0.799
OmyR33562Pearse	5	0.021	0.026	0.019	0	0.043	0.957	2	1
Omy_u09-61.043	5	0.000	0.000	0.000	0.000	0.000	1.000	2	No score, fixed
Omy_RAD30392-17	5	0.402	0.385	0.410	0.214	0.376	0.393	2	0.699
Omy_SECC22b-88	5	0.009	0.000	0.013	0	0.017	0.983	2	0.551
OmyR40319Pearse	5	0.004	0.000	0.006	0	0.009	0.991	2	1
Omy_ndk152	12	0.177	0.244	0.154	0.017	0.333	0.632	NA	0.13444
Omy_IL6320	14	0.177	0.205	0.173	0.060	0.248	0.692	NA	0.62277
Omy_LDHB2_i6	21	0.115	0.115	0.115	0.017	0.205	0.761	NA	0.83429

Results for the 12 markers in the inversion complex of chromosome Omy05 (Pearse et al., 2019) revealed that several were nearly fixed with very low variation in our samples, and others had no significant differences in allele frequencies (**Table 2**). The Omy_u09-61.043 marker was fixed for the resident or derived allele, but this was not the case for all markers in this inversion (**Table 2**). Of the three candidate markers from Narum et al. (2011), two had similar allele frequencies between resident and anadromous groups, and none were significant after a FDR correction (Benjamini and Hochberg

1995). However, the marker with the lowest *p*-value (p = 0.134; Omy_ndk-152; **Table 2**) had an allele frequency difference of 9% between resident and anadromous groups and is located at position 59,538,315 on chromosome Omy12. This is near the region on chromosome Omy12 that was found to be significant in the whole-genome analyses.

4 DISCUSSION

In this study, we conducted multiple analyses to test for genomic regions of high differentiation between the resident and anadromous samples and examine consistent signals that would identify candidate genes associated with this life history trait. Tracking migratory behavior of PIT-tagged juveniles provided resident vs. anadromous (outmigration) phenotypes for genomic analyses, and distinct patterns in migration behavior were observed by each sex. The natural origin *O. mykiss* captured in the Klickitat River screw trap in 2018, 2019, and 2021 were classified as smolts due to distinct physical characteristics at rates of 99.5% (n = 3,615), 99.4% (n = 3,475), and 99.7% (n = 2,352), respectively. Outmigrating individuals detected in the Columbia River after leaving the Klickitat River were most likely steelhead smolts, but downstream migrating resident rainbow trout cannot be completely ruled out. Females rarely remained as residents in contrast to males despite collection from the same stream reaches, which is consistent with the differences in migratory behavior among sexes that has been generally observed in previous studies of this species (Fleming and Reynolds 2004; Quinn et al., 2011). The lack of female resident fish limited genomic analyses to largely focus on variation in male migratory phenotypes, but results revealed a candidate region of minor effect on chromosome Omy12.

Analyses with several approaches were considered necessary to account for limitations of each method and increase confidence in potential candidate genes. For example, F_{ST} and sliding window F_{ST} have been demonstrated to be susceptible to false positives or skewed signal to noise (Fariello et al., 2017), and thus, local score was applied from FET results to improve signal to noise in this high-density genome scan. The RF method was expected to be more effective to detect polygenetic signals of minor effect genes across the genome, especially for natural populations with low linkage disequilibrium (LD) (Holliday et al., 2012). However, analyses from this study yielded highest confidence from tests of sliding window F_{ST} and significance based on FET, whereas other results appeared to suffer from high levels of false positives.

Analyses provided highest confidence for a genomic region of minor effect on chromosome Omy12 for anadromy in male steelhead in the Klickitat River. We chose to align resequencing data from this study to the highest quality genome assembly that was currently available for the species (Gao et al., 2021). Although karyotypes vary across lineages of *O. mykiss*, the number of chromosome arms is constant (Thorgaard et al., 1983; Phillips et al., 2006; Pearse et al., 2019). The newest Arlee assembly has 32 haploid chromosomes due to fissions at chromosomes Omy04, Omy14, and Omy25, but chromosome Omy12 demonstrates homology between Arlee and Swanson lines (Gao et al., 2021). Thus, we do not expect that results observed at Omy12 in the current study were due to problems in homology with the reference genome assembly. The gene nearest to significant SNPs was CSN6 has been shown to be overexpressed in a variety of cancer specimens (Zhang et al., 2013). Another nearby gene on chromosome Omy12 was the NLRP3, which has previously been observed functioning to restrict bacterial infection in Japanese flounder (Chen et al., 2020). The function of these genes did not bear an obvious biological connection to resident or anadromous behaviors and would require further validation before speculating on their related functions in *O. mykiss*.

Six markers specifically developed from the large inversion region on chromosome Omy05 (Pearse et al., 2019) had nearly zero variation among samples in this study and thus were not informative for analyses. This is consistent with results from Pearse et al. (2019) that observed significant results at chromosome Omy05 for coastal populations from the southern portion of the range but were not informative in other portions of the species' range. Additional markers that map to the same region on chromosome Omy05 that were developed in previous restriction site-associated DNA studies (Micheletti and Narum 2018) were variable but not significantly different between the two life history types in the Klickitat River. The difference in variation from markers in the chromosome Omy05 region was unexpected because there is extensive LD across the entire double inversion (Pearse et al., 2019), but this may reflect near fixation of derived alleles followed by subsequent recombination in this population that is part of the distinct inland lineage of *O. mykiss* (e.g., Collins et al., 2020). Previous candidate markers for anadromy in *O. mykiss* from the Klickitat River (Narum et al., 2011) were also not significant, but it was notable that one marker (Omy_ndk-154) was relatively near the same region on chromosome Omy12 identified in the whole-genome scan and had modest difference in allele frequencies between life history types. However, this marker may be too distant (~37 Mb) from the candidate region on chromosome Omy12 to be informative to distinguish life history variation. Thus, there was no evidence in this study population for association with Omy05 migratory behavior as has been shown in a previous study (Pearse et al., 2019). When the results of this study were compared to similar studies (Narum et al., 2011; Hecht et al., 2012; Hale et al., 2013; Hecht et al., 2013; Campbell et al., 2021), a significant association with anadromy was found on chromosome Omy12 in all cases, but these studies have consistently suggested polygenic association located throughout the genome.

Overall, analyses conducted on the *O. mykiss* samples collected from the Klickitat River for this study detected loci significantly associated with the trait of anadromy that may represent candidate genes of minor effect within this *O. mykiss* population. The significant region on chromosome Omy12 provided the strongest support; however, this region contains multiple candidate genes with uncertain biological relevance to resident versus anadromous life history types. Thus, candidate genes of major effect for anadromy remain elusive and results continue to support the hypothesis that the genetic basis for life history types in this lineage of *O. mykiss* is highly polygenic with minor effect (e.g., Nichols et al., 2008; Narum et al., 2011; Hecht

et al., 2012; Hale et al., 2013; Hecht et al., 2013; Campbell et al., 2021). Previous studies evaluated gene expression associated with anadromy at various stages of development of anadromous steelhead and resident rainbow trout and uncovered differentially expressed genes throughout the genome, including many on chromosome Omy12 (McKinney et al., 2015; Hale et al., 2016). In addition, Baerwald et al. (2016) evaluated differentially methylated regions (DMRs) to assess the role of epigenetics on smoltification in *O. mykiss* and identified one DMR associated with smoltification on chromosome Omy12. It is also likely that the main drivers of anadromy in this lineage of *O. mykiss* may be more strongly related to environmental rearing such as epigenetic modification, gene expression, and phenotypic plasticity. Future studies into these various mechanisms regulating anadromy are needed to further understand variation in the life history of this protected species.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: PRJNA771561.

ETHICS STATEMENT

The animal study was reviewed and approved by Monitoring; Methods (monitoringresources.org).

AUTHOR CONTRIBUTIONS

SN and JZ directed the study. SN, JZ, and NR designed the study. EC analyzed the data. All authors interpreted the results and wrote the manuscript.

FUNDING

Funding for this project was provided by the Bonneville Power Administration through grant number 199506335 to Yakima Nation.

ACKNOWLEDGMENTS

We thank YKFP for providing samples (particularly, Kory Kuhn, Sandy Pinkham, Rod Begay, and Dean Antone) and the CRITFC laboratory staff involved in sample processing (particularly, Amber Gonzalez, Megan Moore, Lori Maxwell, Stephanie Harmon, Joe Gasper, and Kim Vertacnik).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.795850/full#supplementary-material

REFERENCES

Baerwald, M. R., Meek, M. H., Stephens, M. R., Nagarajan, R. P., Goodbla, A. M., Tomalty, K. M. H., et al. (2016). Migration-related Phenotypic Divergence Is Associated with Epigenetic Modifications in Rainbow trout. *Mol. Ecol.* 25 (8), 1785–1800. doi:10.1111/mec.13231

Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., et al. (2017). Sex Matters in Massive Parallel Sequencing: Evidence for Biases in Genetic Parameter Estimation and Investigation of Sex Determination Systems. *Mol. Ecol.* 26 (24), 6767–6783. doi:10.1111/mec.14217

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Blankenship, S. M., Campbell, M. R., Hess, J. E., Hess, M. A., Kassler, T. W., Kozfkay, C. C., et al. (2011). Major Lineages and Metapopulations in Columbia RiverOncorhynchus mykissAre Structured by Dynamic Landscape Features and Environments. *Trans. Am. Fish. Soc.* 140 (3), 665–684. doi:10.1080/00028487.2011.584487

Brieuc, M. S. O., Waters, C. D., Drinan, D. P., and Naish, K. A. (2018). A Practical Introduction to Random Forest for Genetic Association Studies in Ecology and Evolution. *Mol. Ecol. Resour.* 18 (4), 755–766. doi:10.1111/1755-0998.12773

Brunelli, J. P., Wertzler, K. J., Sundin, K., and Thorgaard, G. H. (2008). Y-specific Sequences and Polymorphisms in Rainbow trout and Chinook salmon. *Genome* 51 (9), 739–748. doi:10.1139/G08-060

Bushnell, B. (2016). *BBMap Short Read Aligner*. Berkeley, California: University of California http://sourceforge.net/projects/bbmap.

Campbell, M. A., Anderson, E. C., Garza, J. C., and Pearse, D. E. (2021). Polygenic Basis and the Role of Genome Duplication in Adaptation to Similar Selective Environments. *J. Hered.* 112 (7), 614–625. doi:10.1093/jhered/esab049

Campbell, N. R., Harmon, S. A., and Narum, S. R. (2015). Genotyping-in-Thousands by Sequencing (GT-seq): A Cost Effective SNP Genotyping Method Based on Custom Amplicon Sequencing. *Mol. Ecol. Resour.* 15 (4), 855–867. doi:10.1111/1755-0998.12357

Chen, H., Ding, S., Tan, J., Yang, D., Zhang, Y., and Liu, Q. (2020). Characterization of the Japanese Flounder NLRP3 Inflammasome in Restricting *Edwardsiella Piscicida* Colonization *In Vivo*. *Fish Shellfish Immunol.* 103, 169–180. doi:10.1016/j.fsi.2020.04.063

Christie, M. R., Marine, M. L., and Blouin, M. S. (2011). Who Are the Missing Parents? Grandparentage Analysis Identifies Multiple Sources of Gene Flow into a Wild Population. *Mol. Ecol.* 20, 1263–1276. doi:10.1111/j.1365-294X.2010.04994.x

Collins, E. E., Hargrove, J. S., Delomas, T. A., and Narum, S. R. (2020). Distribution of Genetic Variation Underlying Adult Migration Timing in Steelhead of the Columbia River basin. *Ecol. Evol.* 1017, 9486–9502. doi:10.1002/ece3.6641

Courter, I. I., Child, D. B., Hobbs, J. A., Garrison, T. M., Glessner, J. J. G., and Duery, S. (2013). Resident Rainbow trout Produce Anadromous Offspring in a Large interior Watershed. *Can. J. Fish. Aquat. Sci.* 70 (5), 701–710. doi:10.1139/cjfas-2012-0457

Crawford, B. (1979). *The Origin and History of the trout Brood Stocks of the Washington Department of Game*. Olympia: Washington State Game Department, Fisheries Research Report.

Evans, S. D., Lindley, D. S., Kock, T. J., Hansen, A. C., Perry, R. W., Zendt, J. S., et al. (2021). *Evaluation of Movement and Survival of Juvenile Steelhead* (Oncorhynchus mykiss) *and Coho salmon* (Oncorhynchus kisutch) *in the Klickitat River, Washington, 2018–2019 (No. 2021-1083)*. Reston, Virginia, U.S.: US Geological Survey.

Fariello, M. I., Boitard, S., Mercier, S., Robelin, D., Faraut, T., Arnould, C., et al. (2017). Accounting for Linkage Disequilibrium in Genome Scans for Selection without Individual Genotypes: the Local Score Approach. *Mol. Ecol.* 26 (14), 3700–3714. doi:10.1111/mec.14141

Faust, G. G., and Hall, I. M. (2014). SAMBLASTER: Fast Duplicate Marking and Structural Variant Read Extraction. *Bioinformatics* 30 (17), 2503–2505. doi:10.1093/bioinformatics/btu314 https://github.com/GregoryFaust/samblaster

Fleming, I. A., and Reynolds, J. D. (2004). "Salmonid Breeding Systems," in *Evolution Illuminated: salmon and Their Relatives*. Editors A. P. Hendry and S. C. Stearns (Oxford, UK: Oxford University Press), 264–294.

Gao, G., Magadan, S., Waldbieser, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., et al. (2021). A Long Reads-Based De-novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow trout. *G3-Genes GenomGenet* 11, 4. doi:10.1093/g3journal/jkab052

Goldstein, B. A., Polley, E. C., and Briggs, F. B. S. (2011). Random Forests for Genetic Association Studies. *Stat. Appl. Genet. Mol.* 10, 35. doi:10.2202/1544-6115.1691

Hale, M. C., McKinney, G. J., Thrower, F. P., and Nichols, K. M. (2016). RNA-seq Reveals Differential Gene Expression in the Brains of Juvenile Resident and Migratory Smolt Rainbow trout (*Oncorhynchus mykiss*). *Comp. Biochem. Physiol. D: Genomics Proteomics* 20, 136–150. doi:10.1016/j.cbd.2016.07.006

Hale, M. C., Thrower, F. P., Berntson, E. A., Miller, M. R., and Nichols, K. M. (2013). Evaluating Adaptive Divergence between Migratory and Nonmigratory Ecotypes of a Salmonid Fish, *Oncorhynchus mykiss*. *G3-genes Genom. Genet.* 3 (8), 1273–1285. doi:10.1534/g3.113.006817

Hecht, B. C., Campbell, N. R., Holecek, D. E., and Narum, S. R. (2013). Genome-wide Association Reveals Genetic Basis for the Propensity to Migrate in Wild Populations of Rainbow and Steelhead trout. *Mol. Ecol.* 22 (11), 3061–3076. doi:10.1111/mec.12082

Hecht, B. C., Thrower, F. P., Hale, M. C., Miller, M. R., and Nichols, K. M. (2012). Genetic Architecture of Migration-Related Traits in Rainbow and Steelhead Trout, *Oncorhynchus mykiss*. *G3-genes Genom. Genet.* 29, 1113–1127. doi:10.1534/g3.112.003137

Holliday, J. A., Wang, T., and Aitken, S. (2012). Predicting Adaptive Phenotypes from Multilocus Genotypes in Sitka spruce (*Picea Sitchensis*) Using Random forest. *G* 2, 1085–1093. doi:10.1534/g3.112.002733

Horn, R. L., Kamphaus, C., Murdoch, K., and Narum, S. R. (2020). Detecting Genomic Variation Underlying Phenotypic Characteristics of Reintroduced Coho salmon (*Oncorhynchus kisutch*). *Conserv Genet.* 21 (6), 1011–1021. doi:10.1007/s10592-020-01307-0

Kofler, R., Pandey, R. V., and Schlotterer, C. (2011). PoPoolation2: Identifying Differentiation between Populations Using Sequencing of Pooled DNA Samples (Pool-Seq). *Bioinformatics* 27 (24), 3435–3436. doi:10.1093/bioinformatics/btr589

Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv [Preprint]. Available at: https://arxiv.org/pdf/1303.3997.pdf (Accessed September 27, 2021 http://github.com/lh3/bwa).

Li, J., Phillips, R. B., Harwood, A. S., Koop, B. F., and Davidson, W. S. (2011). Identification of the Sex Chromosomes of Brown trout (*Salmo trutta*) and Their Comparison with the Corresponding Chromosomes in Atlantic salmon (*Salmo salar*) and Rainbow trout (*Oncorhynchus mykiss*). *Cytogenet. Genome Res.* 133 (1), 25–33. doi:10.1159/000323410

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R. News* 2 (3), 18–22.http://www.stat.Berkeley.edu/users/breiman/

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome Association and Prediction Integrated Tool. *Bioinformatics* 28 (18), 2397–2399. doi:10.1093/bioinformatics/bts444 http://www.maizegenetics.net/GAPI

Martínez, A., Garza, J. C., and Pearse, D. E. (2011). A Microsatellite Genome Screen Identifies Chromosomal Regions under Differential Selection in Steelhead and Rainbow trout. *Trans. Am. Fish. Soc.* 140 (3), 829–842. doi:10.1080/00028487.2011.588094

McDowall, R. M. (1988). *Diadromy in Fishes*. Portland, OR, US: Timber Press.

McKinney, G. J., Hale, M. C., Goetz, G., Gribskov, M., Thrower, F. P., and Nichols, K. M. (2015). Ontogenetic Changes in Embryonic and Brain Gene Expression in Progeny Produced from Migratory and residentOncorhynchus Mykiss. *Mol. Ecol.* 24 (8), 1792–1809. doi:10.1111/mec.13143

Micheletti, S. J., and Narum, S. R. (2018). Utility of Pooled Sequencing for Association Mapping in Nonmodel Organisms. *Mol. Ecol. Resour.* 18 (4), 825–837. doi:10.1111/1755-0998.1278 https://github.com/StevenMicheletti/poolparty.

Moore, J. W., Yeakel, J. D., Peard, D., Lough, J., and Beere, M. (2014). Life-history Diversity and its Importance to Population Stability and Persistence of a Migratory Fish: Steelhead in Two Large North American Watersheds. *J. Anim. Ecol.* 83 (5), 1035–1046. doi:10.1111/1365-2656.12212

Narum, S. R., Powell, M. S., Evenson, R., Sharp, B., and Talbot, A. (2006). Microsatellites Reveal Population Substructure of Klickitat River Native Steelhead and Genetic Divergence from an Introduced Stock. *North Am. J. Fish. Manage.* 26 (1), 147–155. doi:10.1577/M05-055.1

Narum, S. R., Zendt, J. S., Frederiksen, C., Campbell, N., Matala, A., and Sharp, W. R. (2011). Candidate Genetic Markers Associated with Anadromy In Oncorhynchus mykiss of the Klickitat River. *Trans. Am. Fish. Soc.* 140 (3), 843–854. doi:10.1080/00028487.2011.588131

Narum, S. R., Zendt, J. S., Graves, D., and Sharp, W. R. (2008). Influence of Landscape on Resident and Anadromous Life History Types of *Oncorhynchus mykiss*. *Can. J. Fish. Aquat. Sci.* 65 (6), 1013–1023. doi:10.1139/F08-025

Nichols, K. M., Edo, A. F., Wheeler, P. A., and Thorgaard, G. H. (2008). The Genetic Basis of Smoltification-Related Traits in *Oncorhynchus mykiss*. *Genetics* 179 (3), 1559–1575. doi:10.1534/genetics.107.084251

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., et al. (2019). Sex-dependent Dominance Maintains Migration Supergene in Rainbow trout. *Nat. Ecol. Evol.* 3 (12), 1731–1742. doi:10.1038/s41559-019-1044-6

Phillips, R. B., Nichols, K. M., DeKoning, J. J., Morasch, M. R., Keatley, K. A., Rexroad III, C., et al. (2006). Assignment of Rainbow Trout Linkage Groups to Specific Chromosomes. *Genetics* 174 (3), 1661–1670. doi:10.1534/genetics.105.055269

Quinn, T. P., Seamons, T. R., Vøllestad, L. A., and Duffy, E. (2011). Effects of Growth and Reproductive History on the Egg Size–Fecundity Tradeoff in Steelhead. *Trans. Am. Fish. Soc.* 140, 45–51. doi:10.1080/00028487.2010.550244

Rokach, L. (2016). Decision forest: Twenty Years of Research. *Inf. Fusion* 27, 111–125. doi:10.1016/j.inffus.2015.06.005

Sloat, M. R., and Reeves, G. H. (2014). Individual Condition, Standard Metabolic Rate, and Rearing Temperature Influence Steelhead and Rainbow trout (*Oncorhynchus mykiss*) Life Histories. *Can. J. Fish. Aquat. Sci.* 71 (4), 491–501. doi:10.1139/cjfas-2013-0366

Sweet, D., Lorente, M., Valenzuela, A., Lorente, J., and Alvarez, J. C. (1996). Increasing DNA Extraction Yield from Saliva Stains with a Modified Chelex Method. *Forensic Sci. Int.* 83 (3), 167–177. doi:10.1016/s0379-0738(96)02034-8

Thorgaard, G. H. (1983). Chromosomal Differences Among Rainbow Trout Populations. *Copeia*, 650–662. doi:10.2307/1444329

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B., and Tooker, M. E. (2011). Genomic Inbreeding and Relationships Among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 5673–5682. doi:10.3168/jds.2011-4500

Waples, R. S., Zabel, R. W., Scheuerell, M. D., and Sanderson, B. L. (2008). Evolutionary Responses by Native Species to Major Anthropogenic Changes to Their Ecosystems: Pacific salmon in the Columbia River Hydropower System. *Mol. Ecol.* 17, 84–96. doi:10.1111/j.1365-294X.2007.03510.x

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., et al. (2013). The Sexually Dimorphic on the Y-chromosome Gene (sdY) Is a Conserved Male-specific Y-chromosome Sequence in many Salmonids. *Evol. Appl.* 6 (3), 486–496. doi:10.1111/eva.12032

Zhang, S.-N., Pei, D.-S., and Zheng, J.-N. (2013). The COP9 Signalosome Subunit 6 (CSN6): a Potential Oncogene. *Cell div* 8 (1), 1–5. doi:10.1186/1747-1028-8-14

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed Linear Model Approach Adapted for Genome-wide Association Studies. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fevo.2022.827654

Genetic Diversity and Population Differentiation of Naked Carp (*Gymnocypris przewalskii*) Revealed by Cytochrome Oxidase Subunit I and D-Loop

Di-an Fang[1,2], Hui Luo[2], Miao He[2], Chengcheng Mao[1], Zhen Kuang[1], Hongfang Qi[3], Dongpo Xu[1,2], Longfei Tan[2] and Yuandong Li[1]*

[1] Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, China, [2] National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai, China, [3] Rescues Center of Qinghai Lake Naked Carp, Xining, China

The genetic diversity in the naked carp (*Gymnocypris przewalskii*) of China is threatened by climate change, human activities, as well as natural factors, eliciting conservation concerns. To explore the genetic aspects of *G. przewalskii*, the genetic diversity, genetic structure, population differentiation, and historical demography of 566 representative individuals from seven geographically distinct ranges of Qinghai Lake were evaluated by mitochondrial DNA cytochrome oxidase subunit I (*COI*) and D-loop sequences. Estimates of genetic parameters showed that the seven populations of *G. przewalskii* had high levels of haplotype diversity (0.50243–0.94620) and low levels of nucleotide diversity (0.00079–0.00624). Haplotype genealogy indicated there was no obvious phylogenetic pattern between haplotypes. Both markers denoted the absence of population genetic structure [the genetic differentiation coefficient F-statistics (Fst) < 0] and the presence of high genetic flow (*COI: 0.*9731–1.0441; D-loop: 0.9480–1.0398). The mismatch between the distribution and neutrality tests supported the evidence of population expansion, which occurred during the late middle Pleistocene [*COI*: 0.36–0.108 MYA (Million Years Ago); D-loop: 0.497–0.165 MYA]. Furthermore, this work illustrated two simple, reliable, and inexpensive molecular markers for analysis of genetic diversity, while the sensitivity of the mitochondrial D-loop region as a reflection of genetic diversity in *G. przewalskii* is higher than that of the *COI* gene.

Keywords: *Gymnocypris przewalskii*, genetic diversity, mitochondrial DNA, habitat fragmentation, germplasm protection

OPEN ACCESS

Edited by:
Shaokui Yi,
Huzhou University, China

Reviewed by:
Ailan Wang,
Ludong University, China
Van Lun Low,
University of Malaya, Malaysia

***Correspondence:**
Dongpo Xu
xudp@ffrc.cn

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *02 December 2021*
Accepted: *01 March 2022*
Published: *11 April 2022*

Citation:
Fang D-a, Luo H, He M, Mao Cc, Kuang Z, Qi Hf, Xu Dp, Tan Lf and Li Yd (2022) Genetic Diversity and Population Differentiation of Naked Carp (Gymnocypris przewalskii) Revealed by Cytochrome Oxidase Subunit I and D-Loop. Front. Ecol. Evol. 10:827654. doi: 10.3389/fevo.2022.827654

INTRODUCTION

Naked Carp (*Gymnocypris przewalskii*) is a cyprinid fish that mainly inhabits Qinghai Lake of China at an altitude of 3,000–4,000 m (Wu et al., 1964; Tong et al., 2015). This species, the unique dominant population in Qinghai Lake, is an endemic and ecologically important fish in the Northwest Plateau ecosystem (Zhang and Zhang, 1963; Wang and Xie, 2004; Wang et al., 2014). Particularly, *G. przewalskii* migrates into freshwater tributaries connected with Qinghai Lake to spawn from April to July every year. Then, they migrate back to Qinghai Lake after the reproductive behavior. However, much less is known about the specific migratory routes and whether there are

genetic differences between different geographical populations of *G. przewalskii*. Due to overfishing, dam construction, water pollution, and environmental factors like climate, the wild populations of *G. przewalskii* have declined dramatically in size, and the utilization of this resource has far exceeded appropriate limits (He, 2008; Luo, 2015). In response, and to evaluate this resource, the government has developed a series of initiatives, including the artificial restoration of migratory channels and fishery stocking (Shi, 2008). Such a stocking may evaluate resources, but may also result in negative impacts, such as loss of genetic diversity and fitness, even causing genetic risk (Anneville et al., 2015). Thus, additional research is needed on the status of germplasm resources. Currently, research on *G. przewalskii* mainly focused on fingerling breeding technology, otolith marking, gene cloning, expression, and so on (Cao et al., 2009; Feng et al., 2019, 2021; Min et al., 2020; Wu et al., 2020; Jiang et al., 2021). Unfortunately, relatively little is understood about this species' genetic characteristics (Shi, 2008).

China's wildlife species account for approximately 11% of the world, so strengthening China's wildlife conservation is of great significance to global biodiversity (Gong et al., 2020). Genetic characterizations of natural populations are key to providing a point of surveillance and conservation (Liu, 2004; Zhang et al., 2015; Yan et al., 2020). Knowledge of the genetic variation in natural populations and the implementation of an efficient management plan based on genetic features are important measures for its maintenance and recovery (Cross, 2015; Rômulo et al., 2018). Nevertheless, limited effort has been made to monitor the genetic diversity of *G. przewalskii* on wild populations in China. Therefore, it is necessary to assess the genetic diversity of *G. przewalskii* to understand the conservation status and develop effective protection measures.

According to present knowledge, early studies performed a restriction fragment length polymorphism analysis in *G. przewalskii* and reported high levels of genetic diversity (Zhao et al., 2001; Qi, 2002; Xu et al., 2003). Similarly, based on the mitochondrial D-loop region, Zhang et al. (2013) found that *G. przewalskii* possessed high levels of genetic diversity and a certain degree of differentiation among *G. przewalskii* in different geographical populations (Qinghai Lake, Crooke Lake, Ganzi River, and Cao Dalian). The analyses of amplified fragment length polymorphism and simple sequence repeat further demonstrated a high level of genetic diversity, which may be related to the origin and evolution of *G. przewalskii* and the specificity of the plateau environment (Jiang et al., 2009; Wang et al., 2015). Nonetheless, some researchers reported that there existed quite a weak genetic differentiation between the Heima River, Buha River, and Shaliu River populations and quite a strong gene exchange among the three populations, simultaneously (Zhang et al., 2005; Chen et al., 2006). This point of view was also supported by Zhao et al. (2006) who found evidence for the absence of extensive population structure and a low level of diversity. These issues addressed in previous studies obtained discrepant or even paradoxical results. Hence, it is urgently needed to have a better understanding of the genetic characteristics of *G. przewalskii*.

Fish mitochondrial DNA has a simple structure, small molecular weight, high conservation, and maternal inheritance. The genetic diversity of fish can be more reliably analyzed by comparing base sequence composition and analyzing mutation sites (Perez-Enriquez et al., 1999). Mitochondrial cytochrome oxidase subunit I (*COI*) is a relatively conservative gene sequence with a moderate evolution rate. It is one of the most thoroughly studied mitochondrial genes and is ideal as a molecular marker (Sari et al., 2015). At the same time, the D-loop control region has the characteristics of rapid evolution and more genetic variation because of no coding pressure, which is suitable for detecting the genetic diversity within and between populations (Gatt et al., 2000). In this study, the genetic diversity and population differentiation of *G. przewalskii* from 7 different locations in Qinghai Lake were evaluated based on the *COI* gene and D-loop control region to provide a theoretical basis for their resource protection.

MATERIALS AND METHODS

Materials

A total of 566 *G. przewalskii* individuals were collected randomly from the Buha River Estuary (BHE), Yilang Jian (YLJ), Heima River (HMR), Buha River (BHR), Quanji River (QJR), Shaliu River (SLR), and Haergai River (HeGR) from 2019 to 2020 (**Table 1** and **Figure 1**), which were representative. The detailed geographical positions can be seen in **Figure 1**. These five rivers are the largest tributaries connected with Qinghai Lake, where *G. przewalskii* migrate to spawn annually, and the YLJ area is located in the Qinghai Lake area, where *G. przewalskii* returns after spawning. After the morphological identification (Zhao et al., 2005), the caudal fins of each individual were collected and preserved in anhydrous ethanol at room temperature for later use.

DNA Extraction, Amplification, and Sequencing of the Mitochondrial Fragments

No more than 30 mg of each fin were used for genomic DNA extraction using a genomic DNA extraction kit (TIANamp

TABLE 1 | Details for sampling sites and quantities of *Gymnocypris przewalskii*.

Locality	GPS	Abbreviation	Numbers	Tissues	Date
Buha River	99.74°E, 37.04°N	BHR	81	Caudal fin	2019-05
Heima River	99.77°E, 36.73°N	HMR	80	Caudal fin	2019-07
Quanji River	99.89°E, 37.24°N	QJR	80	Caudal fin	2020-07
Shaliu River	100.18°E, 37.24°N	SLR	81	Caudal fin	2019-06
Haergai River	100.48°E, 37.22°N	HeGR	83	Caudal fin	2020-08
Buha River Mouth	99.88°E, 36.95°N	BHE	82	Caudal fin	2019-08
Yilangjian	100.40°E, 36.70°N	YLJ	79	Caudal fin	2020-06
Total			566		

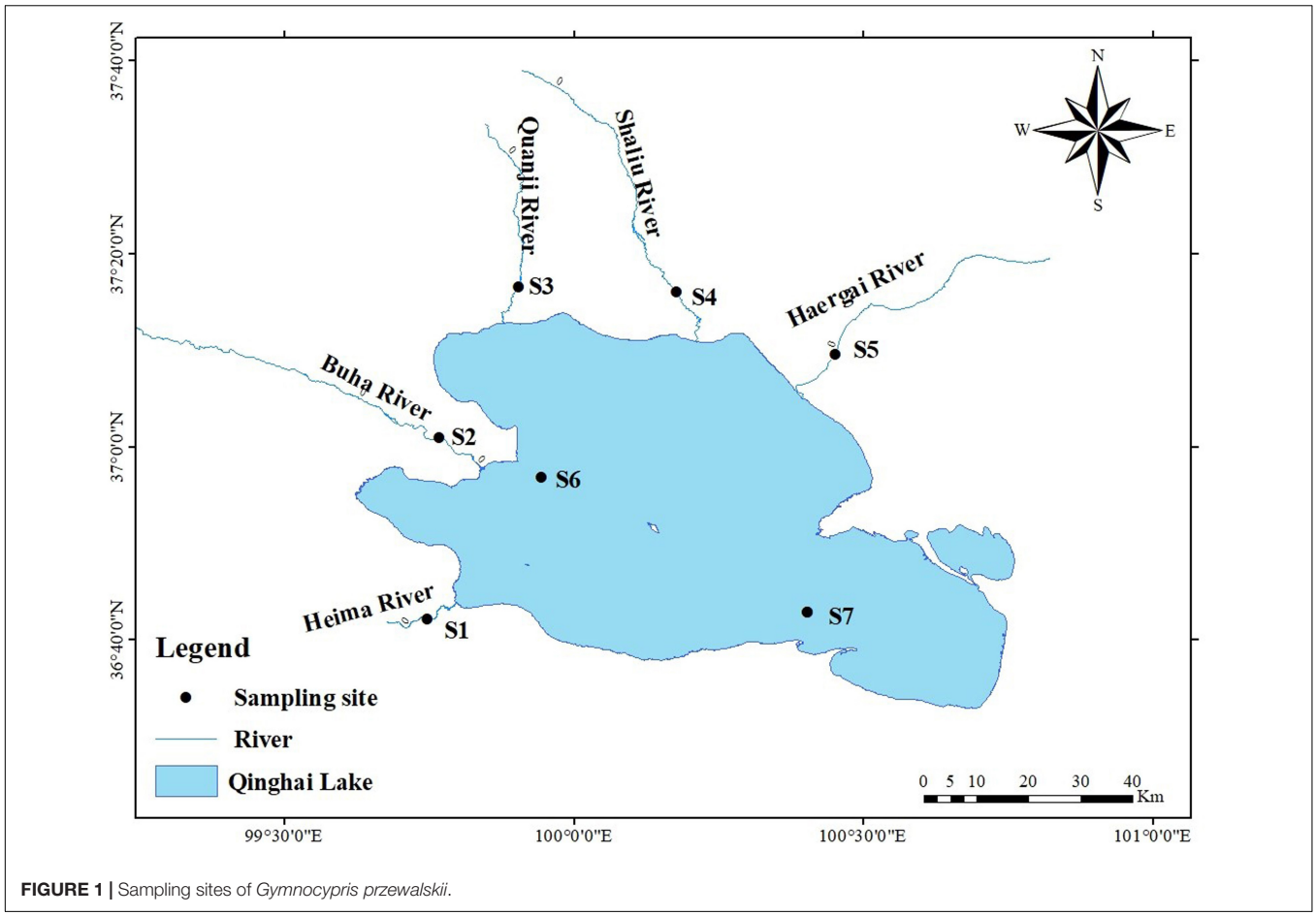


FIGURE 1 | Sampling sites of *Gymnocypris przewalskii*.

Marine Animals DNA Kit, TIANGEN) in the laboratory. The extracted DNA was verified by electrophoresis in 1% agarose gel and then stored at -20°C until use. The fragments of the *COI* gene were amplified with primers L5956-*COI* (5′-CACAAAGACATTGGCACCCT-3′) and H6855-*COI* (5′-AGTCAGCTGAAKCTTTTAC-3′) (Miya and Nishida, 2000). The fragments of the D-loop control region were amplified with primers D-loop-F5 (5′-GGGATATGTCATCCTTTATGG-3′) and D-loop-R5 (5′-GGGTTTGACAAGAATAACAGG-3′) (Zhang et al., 2015). PCR analysis was performed using 25 μl final volume containing 2 μl template DNA, 1.25 μl of upstream and downstream primers, 12 μl of premix, and 8.5 μl of double-distilled H_2O. After an initial 5 min denaturing step at 94°C, 35 cycles of amplification were performed at 94°C for 10 s, annealing for 10 s (*COI* 57°C and D-loop 58°C), elongation at 72°C for 30 s, and a final extension at 72°C for 10 min. The PCR products were verified by 2% agarose gel electrophoresis. The purified PCR products were sequenced at the Shanghai Yixin Biotechnology Company (Shanghai, China).

TABLE 2 | Nucleotide composition of *G. przewalskii* populations by two mitochondrial markers.

Population	A (%)		T (%)		C (%)		G (%)	
	COI	D-loop	*COI*	D-loop	*COI*	D-loop	*COI*	D-loop
HMR	24.44	31.12	32.89	31.71	24.11	22.35	18.56	14.83
BHR	24.43	31.12	32.89	31.72	24.12	22.34	18.56	14.81
YLJ	24.43	31.13	32.89	31.78	24.11	22.30	18.56	14.80
BHE	24.44	31.11	32.89	31.4	24.12	22.33	18.55	14.82
HeGR	24.43	31.11	32.89	31.76	24.11	22.32	18.57	14.82
SLR	24.43	31.11	32.89	31.74	24.11	22.32	18.57	14.83
QJR	24.44	31.12	32.90	31.70	24.11	22.37	18.55	14.82
Average	24.44	31.12	32.89	31.73	24.11	22.33	18.56	14.82

Data Analysis

Nucleotide sequences obtained were aligned using Clustal X 2.0 with the default settings and manually examined (Larkin et al., 2007). The base composition, genetic distance within and between populations, variable sites, and parsimony-informative sites were determined using the MEGA X (Kumar et al., 2018) software. The phylogenetic neighbor-joining (NJ) tree was built using the Kimura 2-parameters method with MEGA X, with 1,000 bootstrap replicates (Kumar et al., 2018). We compared our *COI* sequences with *Gymnocypris dobula* (accession number HQ198864.1), *Gymnocypris namensis* (accession number HQ198865.1), *Gymnocypris firmispinatus* (accession number MF122270.1), and *Gymnocypris potanini* (accession

number MK716230.1), and compared our D-loop sequences with *G. dobula* (accession number MK886318.1), *G. namensis* (accession number MK886398.1), *Gymnocypris waddellii* (accession number KY461279.1), and *G. potanini* (accession number MG756690.1) that were downloaded from GenBank. We used the *COI* sequence (accession number JX983283.1) and D-loop sequence (accession number KX576711.1) of *Cyprinus carpio* from GenBank as the outgroup. Additionally, the gene flow (Nm) and genetic diversity parameters such as haplotype number (*h*), haplotype diversity (*Hd*), and nucleotide diversity (*pi*) were inferred using the program DnaSP v6.0 (Julio et al., 2017) for downstream analyses. The relationships among the samples were visualized using a haplotype network constructed based on the Network 10.2 software (Bandelt et al., 1999). The analysis of molecular variance (AMOVA) was determined in Arlequin 3.5 (Laurent and Heidi, 2010). Furthermore, the historical population expansion was examined using a neutrality test and the mismatch distribution in DnaSP v6.0. For the neutrality test, the values of Tajima's D and Fu's *FS* were calculated (Tajima, 1989; Fu and Li, 1993). The timing of possible population expansions (t) was calculated from the relationship $t = T\tau /2 \mu k$, where T was the generation time, τ was the mode of the mismatch distribution, μ was the mutation rate per nucleotide, and k was the number of nucleotides of the analyzed fragment (Rogers and Harpending, 1992). The time of initial maturity for *G. przewalskii* was estimated to be 3–4 years. Here, we used 4 years as the generation time (Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2003). We used a range of mutation rates (COI: 1–3% per million years, D-loop: 3–10% per million years) (Donaldson and Wilson, 1999; Wang et al., 2013) and the value of τ was calculated using Arlequin 3.5 (Laurent and Heidi, 2010).

RESULTS

Nucleotide Composition

After submitting the company's sequencing results to the NCBI database and editing for calibration, among the 566 individuals of seven regional populations of *G. przewalskii*, the partial sequences of the *COI* gene were 725 base pairs (bp), and the length of the D-loop region was 611 bp. The A+T content was rich (57.33 and 62.85%, respectively) in the nucleotide composition of the *COI* and D-loop sequences, and the D-loop region had the higher A+T content compared to that in *COI* (**Table 2**).

For the *COI* gene, 17 haplotypes were identified. Among them, only four haplotypes (Hap 1, Hap 2, Hap 5, and Hap 6) were found in seven populations, and Hap1 was the most frequent one (**Table 3** and **Figure 2**). For the D-loop region, 66 haplotypes were identified. Among them, eleven haplotypes were found in seven populations, and Hap1 and Hap4 were the most frequent ones (**Table 3** and **Figure 3**).

Genetic Diversity

A total of 15 and 55 variable sites, including 8 and 38 parsimony informative sites, were observed in the *COI* and D-loop region sequences, respectively. The *h*, *Hd,* and *Pi* of the *COI* gene in different populations ranged from 5 to 10, 0.50243 to 0.67184, and 0.00079 to 0.00118, respectively (**Table 3**). Among them, the QJR population had the highest *h, Hd, and Pi* (10, 0.67184, and 0.00118, respectively) (**Table 3**).

Additionally, the *h*, *Hd,* and *Pi* of the D-loop region in different populations ranged from 23 to 31, 0.91790 to 0.94620, and 0.00509 to 0.00632, respectively (**Table 3**). A closer inspection of the table showed the *h*, *Hd,* and *Pi,* for each population based on the D-loop region, were higher than those based on the *COI* gene. Besides, the Nm values based on the *COI* and D-loop sequence ranged from 0.9731 to 1.0441 and 0.9480 to 1.0398, respectively (**Table 4**). The gene flow between the HMR population and the QJR population was minimal based on two markers.

Population Differentiation and Structure

The construction of the *COI* haplotype network (**Figure 2**) was based on 17 different haplotypes. Some haplotypes were shared by different groups, while others were exclusive to one group. For example, Hap 1, Hap 2, Hap 5, and Hap 6 were shared by all seven populations. On the contrary, Hap 3 was exclusive to the BHR population, Hap 14 was exclusive to the QJR population, and Hap 12 was exclusive to the HMR population. Many haplotypes formed clusters around Hap 1, and Hap 1 appeared the most frequently. Other haplotypes, except Hap 3, Hap 11, Hap 12, and Hap 14, were all differentiated from Hap 1. It is inferred that Hap 1 was the initial haplotype, and there was no obvious phylogenetic pattern between haplotypes.

The NJ tree was based on the *COI* marker, and the same samples gave similar results (**Figure 4**). The results showed that the haplotypes were divided into two branches with Hap 1 as the axis and the other three small branches outside. It should be noticed that *G. dobula* and *G. namensis* had the closest relation with *G. przewalskii*. In addition, the genetic distance between the haplotypes of *G. przewalskii* ranged from 0.00139 to 0.00698 based on the *COI* gene. The genetic distances between Hap 1, Hap 2, Hap 5, and other haplotypes were relatively small, and the genetic distances between Hap 3, Hap 14, and Hap 12 were the farthest.

Turning now to the experimental evidence on the D-loop region, the construction of the D-loop haplotype network (**Figure 3**) was based on 66 different haplotypes. The 66 haplotypes were divided into 4 clusters: Group 1 centered around Hap 17, Group 2 centered around Hap 8, Group 3 centered around Hap 1, and Hap 57 as a single group. In addition, Hap 61, Hap 62, and Hap 63 were exclusive to the HMR population, Hap 28 and Hap 31 were exclusive to the GJR population, and Hap 20 was shared by the QJR and SLR population.

A ring for the NJ tree of the D-loop region was used for better presentation. The NJ phylogenetic tree of the D-loop region showed that the lineages of *G. przewalskii* were divided into several single families and were not clustered according to each region. It should be noticed that *G. potanini* and *G. waddellii* had the closest relation with *G. przewalskii* (**Figure 5**). Furthermore, the results of the sequence analysis based on the D-loop region showed the genetic distance that spans between haplotypes was small, ranging from 0 to 0.02831. Among them, the genetic

TABLE 3 | Genetic diversity indices of *G. przewalskii* populations by two mitochondrial markers.

Population	Sample size	N		*h*		*Hd*		*Pi*	
		COI	D-loop	*COI*	D-loop	*COI*	D-loop	*COI*	D-loop
BHR	81	8	29	8	23	0.53179	0.91790	0.00092	0.00558
HMR	80	7	35	8	28	0.55063	0.94620	0.00094	0.00509
QJR	80	9	31	10	26	0.67184	0.93259	0.00118	0.00534
SLR	81	8	31	8	25	0.54259	0.93179	0.00091	0.00624
HeGR	83	6	32	7	31	0.53041	0.92418	0.00084	0.00598
BHE	82	5	27	6	24	0.54682	0.92472	0.00088	0.00600
YLJ	79	4	34	5	25	0.50243	0.92308	0.00079	0.00632
Total	566	15	55	17	66	0.55221	0.92822	0.00092	0.00578

N, number of variable sites; h, number of haplotypes; Hd, haplotype diversity; Pi, nucleotide diversity.

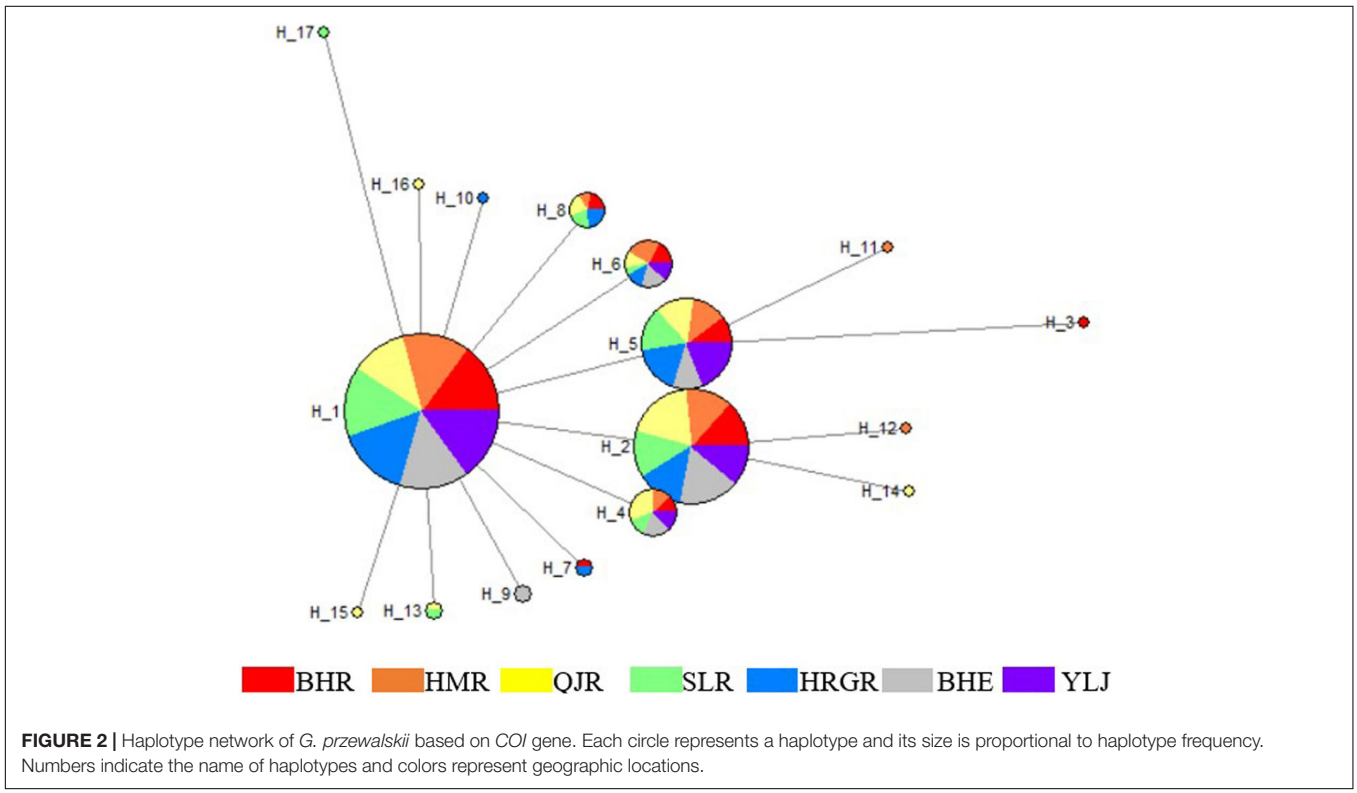


FIGURE 2 | Haplotype network of *G. przewalskii* based on *COI* gene. Each circle represents a haplotype and its size is proportional to haplotype frequency. Numbers indicate the name of haplotypes and colors represent geographic locations.

distances between Hap 31 and Hap 61, Hap 28 and Hap 62, and Hap 20 and Hap 63 were 0. The genetic distances between Hap 57 and Hap 53 as well as Hap 57 and Hap 13 were both the maximum value of 0.02831. Hap 57 and Hap 53 were exclusive to the YLJ population, while Hap 13 was shared by the HeGR, SLR, and QJR populations.

The AMOVA revealed that most of the molecular variance came from the within-population based on the two markers (**Table 5**). The total variability observed among populations in the *COI* and D-loop region sequences were –0.31 and –0.24, respectively, whereas 100.31 and 100.24% of the variation were found within populations, respectively. Moreover, the genetic differentiation coefficient F-statistics (Fst) for both the mt DNA molecular markers were negative values, which indicated that the degree of genetic differentiation of *G. przewalskii* was not high and there was no obvious genetic differentiation between *G. przewalskii* populations (**Table 5**).

Historical Demography

The neutrality tests (**Table 6**) and mismatch distribution (**Figure 6**) based on the *COI* gene supported the hypothesis of population expansion of *G. przewalskii*. The Tajima's D-test and the Fu's FS-test of *COI* gene showed negative values that reached a significant level (total: $P < 0.05$, **Table 6**). Also, a unimodal pattern was observed for all populations except the YLJ population in the mismatch distribution analyses based on the *COI* gene (**Supplementary Figure 1**). However, both the Tajima's D-test and the Fu's FS-test of the D-loop region were negative, with the former not reaching a significant level while the latter did (Tajima's D p-value > 0.05, Fu's FS p-value < 0.05, **Table 6**).

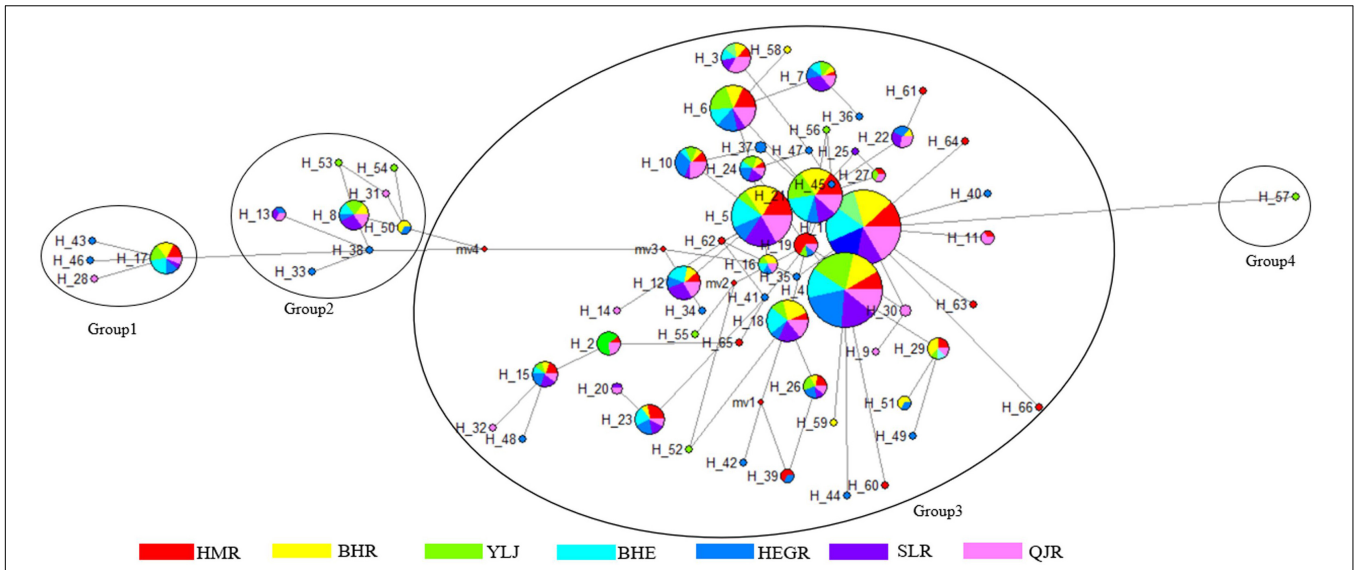


FIGURE 3 | Haplotype network of D-loop region in *G. przewalskii*. Each circle represents a haplotype and its size is proportional to the haplotype frequency. The numbers indicate the name of the haplotypes and the colors represent geographic locations.

The bimodal distribution was observed for all populations in the mismatch analyses based on the D-loop region (**Figure 6** and **Supplementary Figure 2**). In terms of the beginning of population expansion, the τ value of 0.781 and 3.038 derived based on *COI* and D-LOOP suggested a time of approximately 0.36–0.108 MYA (Million Years Ago) and 0.497–0.165 MYA, occurring during the late middle Pleistocene.

TABLE 4 | Gene flow (Nm) analysis of *G. przewalskii* populations by two mitochondrial markers.

Population	BHR	HMR	QJR	SLR	HEGR	BHE	YLJ
BHR		1.0441	1.0043	1.0401	1.0352	1.0304	1.0214
HMR	1.0062		1.0135	1.0352	1.0343	1.0355	1.0260
QJR	1.0115	1.0225		1.0051	1.0150	1.0224	0.9731
SLR	1.0313	1.0030	1.0117		1.0406	1.0180	1.0351
HeGR	1.0282	1.0089	1.0053	1.0302		1.0055	1.0360
BHE	1.0394	1.0106	1.0090	1.0398	1.0293		0.9937
YLJ	0.9989	0.9541	0.9480	1.0075	1.0077	1.0045	

Pairwise estimates of Nm based on COI gene (above diagonal) and D-loop region (below diagonal) for seven populations of G. przewalskii.

DISCUSSION

Nucleotide Composition

Our finding that the content of A+T (57.33%) was higher than that of C+G (42.67%) of *COI* gene in *G. przewalskii* is consistent with the previous research in *Aspiorhynchus laticeps* (A+T = 54.9%), *Schizothorax biddulphi* (A+T = 53%, 53.1%), and *Schizothorax irregularis* (A+T = 53.2%) (Yang et al., 2011, 2013). This confirms our findings and indicates that the content of A+T was higher than that of C+G, which was common in the sequence composition of the *COI* gene of *schizothoracids*. In addition, our result based on the D-loop region reflects that of Zhang et al. (2013) who also found that the content of A+T (63.1%) was significantly higher than the content of C+G (36.9%) in *G. przewalskii.*

Comparison of Genetic Diversity Based on Cytochrome Oxidase Subunit I Gene and D-Loop Region

One interesting finding is that the *h* (17 and 66, respectively) and *Hd* (0.55221 and 0.92822, respectively) among populations of *G. przewalskii* differed by the *COI* and D-loop regions. The difference may be attributed to the evolutionary rate of different molecular markers (Wu, 2017; Yan et al., 2020). The mutation rate and evolution rate in the D-loop region, as no selection pressure for coding, are higher than that in the *COI* gene. Clearly, the current results showed both the mitochondrial *COI* gene and D-loop region markers can be very helpful in analyzing the genetic variation of *G. przewalskii*. In addition, the *Pi* of mitochondrial DNA is one of the crucial indicators to measure the genetic diversity of the population. The larger the *Pi* value, the higher the polymorphism of the population. What emerges from the results reported here is that both the *h* and *Pi* of *G. przewalskii* from the *COI* gene were lower than those from the D-loop region, indicating that the sensitivity of the D-loop control region to reflect the *genetic* diversity of *G. przewalskii* is higher than that of the *COI* gene. It also confirmed the view that the evolution rate of the D-loop control region is higher than that of the *COI* gene.

Moreover, according to Millar et al. (1991), when Nm > 1, the gene flow between populations is considered high and genetic similarity is high. Except for the weak genetic flow between the YLJ and QJR, BHE, BHR, HMR populations, respectively, all other populations showed a high genetic flow (Nm > 1),

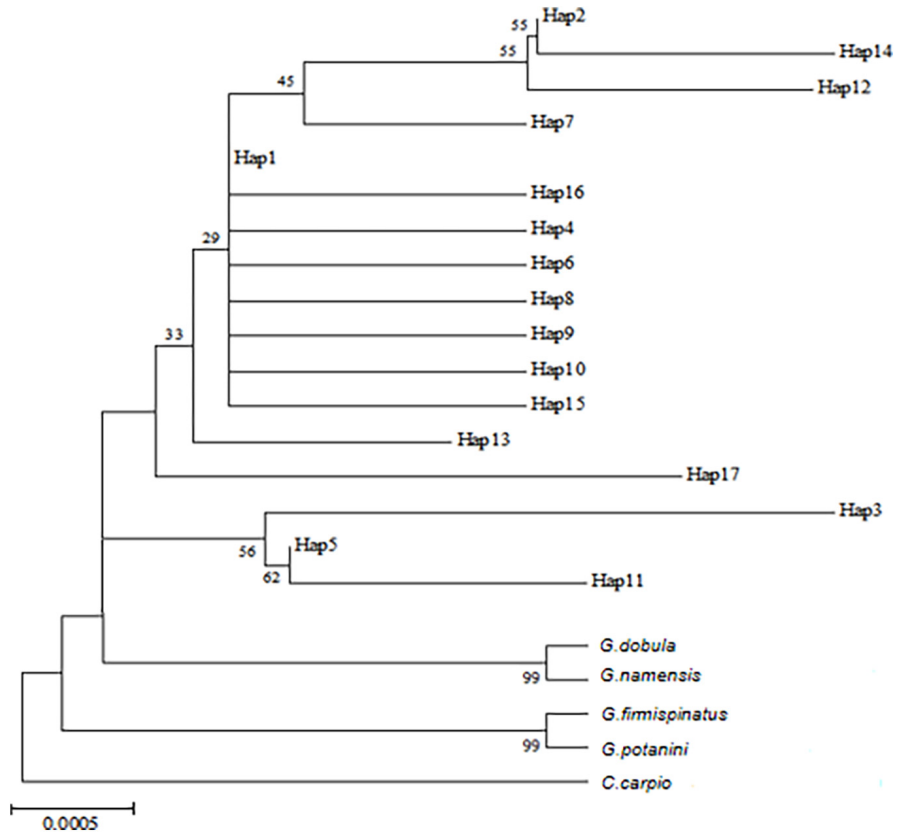


FIGURE 4 | NJ tree of *COI* gene of *G. przewalskii* based on Kimura-2-parameter model. Numbers at the nodes represent bootstrap values.

indicating high genetic similarity among the populations, thus, causing a low level of genetic differentiation, as has been reported previously for *Macropodus ocellatus* in the Yellow River and Yangtze River (Zhang et al., 2019). Next, the gene flow among YLJ and the other populations was low based on two markers. This might be because the YLJ population is located in the southeast of Qinghai Lake and no rivers are entering the lake nearby, which makes it more difficult to migrate to other rivers. Therefore, the YLJ population may not participate in reproductive migration.

Population Differentiation

Another momentous finding is that the genetic variation of *G. przewalskii* primarily occurred within the population (Fst *COI* = −0.00308; Fst D-loop = −0.00240) based on the above two molecular markers. The result may be associated with the reproductive migration of *G. przewalskii*. The migration route of *G. przewalskii* may be varied every year. Such a migration process has greatly increased the gene exchange of the entire lake area. In this study, the result based on the D-loop region was similar to the results of the previous study (Fst = 0.01926) in the HMR, BHR, and SLR (Chen et al., 2006). The previous studies estimated from the D-loop sequence demonstrated that there was little genetic differentiation among the three reproductive populations of *G. przewalskii*, and genetic variation mainly existed within the population (Chen et al., 2006). Likewise, it was previously reported that the gene exchange of *G. przewalskii* was so frequent that no obvious population structure was found by sequencing the mitochondrial cytochrome b (*Cytb*) gene (Zhao et al., 2006). However, using an evaluation based on random amplified polymorphic DNA (RAPD) markers, Zhang et al. (2005) showed that there was little genetic differentiation (Fst = 0.1070) between three *G. przewalskii* populations. This differs slightly from the findings presented here probably due to the different molecular marker methods.

Genetic Structure

Estimates of population genetic parameters showed that *G. przewalskii* had high levels of haplotype diversity (0.55221 and 0.92822, respectively) and low levels of nucleotide diversity (0.00092 and 0.00578, respectively) based on the *COI* gene and D-loop region. Historically, the values defining the level of haplotype diversity and nucleotide diversity by Grant were 0.5 and 0.005, respectively (Grant and Bowen, 1998). Hence, the structure of *G. przewalskii* in this study was more inclined to the type of high haplotype and low nucleotide diversity as previously reported (Zhang et al., 2013). Similarly, previous research recorded the haplotype and nucleotide diversities from the mitochondrial DNA *Cytb* gene of populations of *G. przewalskii* (0.8042 ± 0.0424 and 0.00201 ± 0.00124, respectively), *G. eckloni* of the Yellow River (0.8720 ± 0.0205 and 0.00485 ± 0.00249,

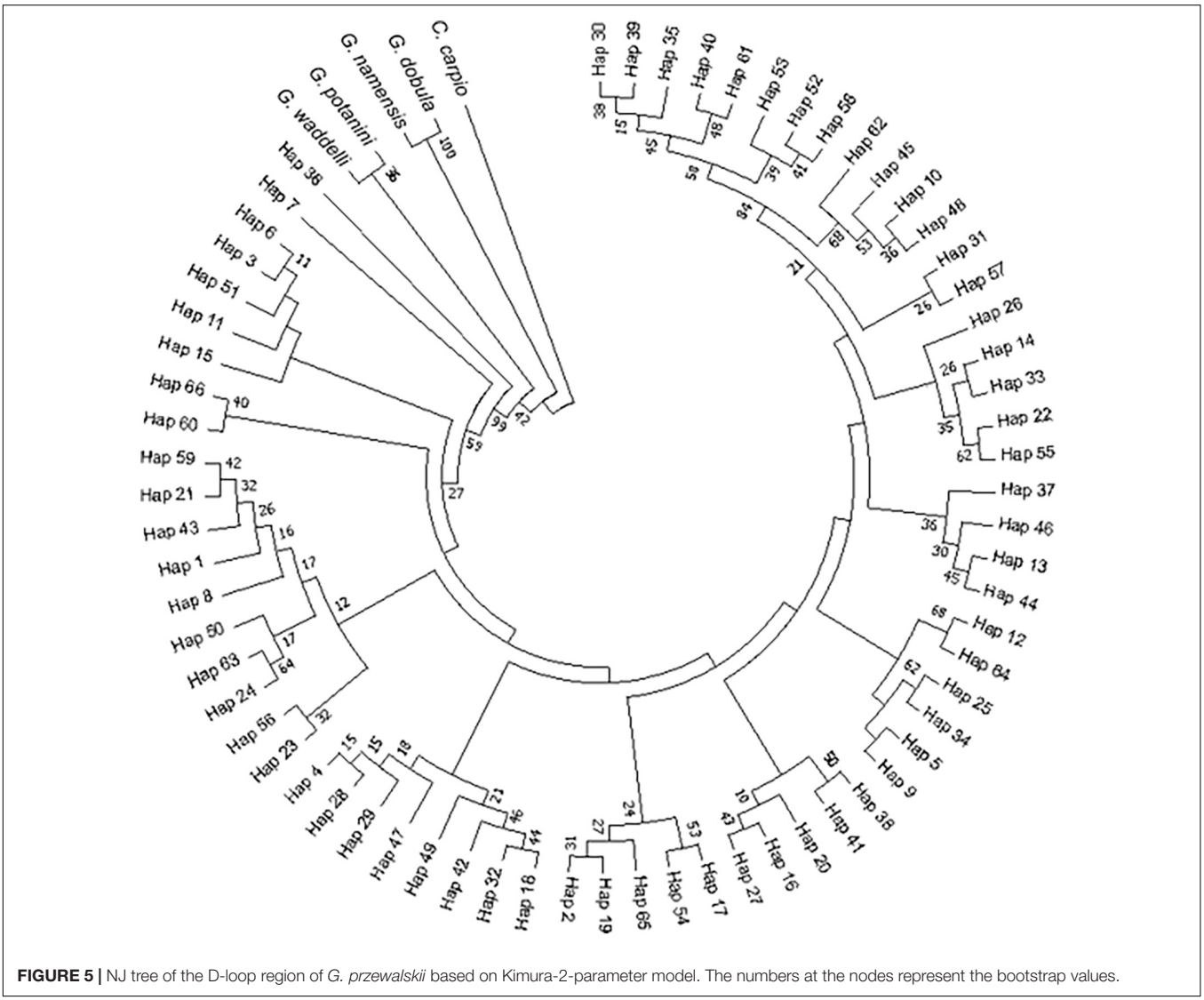


FIGURE 5 | NJ tree of the D-loop region of *G. przewalskii* based on Kimura-2-parameter model. The numbers at the nodes represent the bootstrap values.

TABLE 5 | Analysis of molecular variance (AMOVA) for the *G. przewalskii* populations.

Source of variation	*df*		Sum of squares		Variance components		Percentage of variation (%)	
	COI	D-loop	*COI*	D-loop	*COI*	D-loop	*COI*	D-loop
Among populations	6	6	1.249	9.598	–0.00085 Va	–0.00476 Va	–0.31	–0.24
Within populations	559	559	154.751	1109.336	0.27684 Vb	1.98450 Vb	100.31	100.24
Total	565	565	156.000	1118.935	0.27599	1.97974		

df, degrees of freedom; Fst (COI) = -0.00308; Fst (D-loop) = -0.00240; (P > 0.05).

respectively), and *G. eckloni* of the Geerm River (0.5771 ± 0.1182 and 0.00096 ± 0.00073, respectively) (Zhao et al., 2005).

Generally, a large population, different environments, and living habits suitable for the rapid growth of the population are required for the persistence of high haplotype diversity of the population in nature (Grant and Bowen, 1998). As it happens, these situations are present on *G. przewalskii*. First, through the measures implemented by the Qinghai Provincial Government, such as embargo fishing and stock enhancement of *G. przewalskii*, the resource of *G. przewalskii* has been restored to 29% of the original reserve. The resource has increased from 26,000 tons in 2002 to 50,000 tons in 2014 and 93,000 tons in 2019. Second, the ecological environments of lakes and rivers are different, which can better provide diverse conditions for the feeding, reproduction, and over-wintering of *G. przewalskii*. Finally, *G. przewalskii* will undergo reproductive migration from April to July every year. They migrate from the Qinghai Lake into the various rivers connected with the Qinghai Lake. After

the reproductive behavior is over, they swim back to the Qinghai Lake. The routine habits have led to a yearly recovery of *G. przewalskii* year by year. Therefore, it is speculated that the type of high haplotype and low nucleotide diversity of *G. przewalskii* comes from these conditions.

Population History

The Tajima's D-tests and the Fu's Fs-test were not uniform for the D-loop region, which has a slightly faster evolutionary rate and cannot accurately analyze the population dynamics of *G. przewalskii* combined with the nucleotide mismatch distribution. Therefore, the results of the *COI* gene analysis were applied to determine whether there was population expansion. In short, evidence from the neutrality tests and mismatch distribution revealed that *G. przewalskii* may have experienced population expansion (Tajima, 1989; Fu and Li, 1993). This may be the primary reason why little population differentiation was detected. A similar result was obtained for *G. dobula* from three localities (Pali, Lasa, and Yanghu) in the Tibetan plateau (Chan et al., 2016). Simultaneously, these outcomes were in accordance with the pattern of high haplotype diversity and low nucleotide diversity observed, with the "star-like" shape of the haplotype network in this study, and with the variation of resources in *G. przewalskii*.

To our knowledge, the complicated climatic changes as the Tibetan plateau uplift play a crucial role in the demographic history of fish species (Gang et al., 2009; Chan et al., 2016). In the Tibetan plateau, the largest glacial retreat has been occurring since 0.17 MYA (Zheng et al., 2002). Based on the expansion time we studied (0.36–0.108 MYA, 0.497–0.165 MYA), the demographic history could conform with the period of the largest glacial retreat during the Pleistocene. Therefore, we suggest that climatic conditions during the Pleistocene period had a degree of influence on the contemporary distribution of the genetic variation of *G. przewalskii*, as reported in *G. dobula* (Chan et al., 2016). Additionally, the population expansion occurring in the Pleistocene agreed with the previous findings of schizothoracine fishes in the Yellow River. The Yellow River schizothoracine fishes have younger expansion than *G. przewalskii* (Duan et al., 2009).

TABLE 6 | Neutrality test of *G. przewalskii* by two mitochondrial markers.

Population	Gene	Tajima's D		Fu's FS	
		D	*P*	FS	*P*
BHR	*COI*	−1.48136	>0.10	−4.105	0.012
	D-loop	−1.07559	>0.10	−7.549	0.000
HMR	*COI*	−1.26351	>0.10	−3.990	0.013
	D-loop	−1.64363	>0.05	−16.648	0.000
QJR	*COI*	−1.37040	>0.10	−5.309	0.004
	D-loop	−1.30112	>0.10	−11.709	0.000
SLR	*COI*	−1.49289	>0.10	−4.157	0.011
	D-loop	−0.98521	>0.10	−8.352	0.000
HeGR	*COI*	−1.14505	>0.10	−3.206	0.027
	D-loop	−1.13029	>0.10	−16.454	0.000
BHE	*COI*	−0.81948	>0.10	−1.981	0.077
	D-loop	−0.74558	>0.10	−7.680	0.000
YLJ	*COI*	−0.62079	>0.10	−1.286	0.131
	D-loop	−1.23569	>0.10	−8.395	0.000
Total	*COI*	−1.78116	<0.05	−13.289	0.000
	D-loop	−1.44274	>0.10	−32.562	0.000

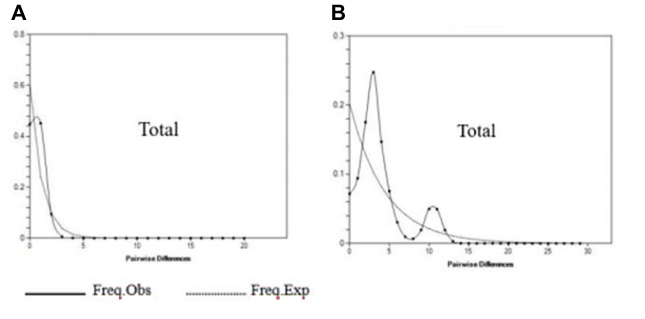


FIGURE 6 | Nucleotide mismatch distribution in *G. przewalskii* of all populations for two markers: **(A)** *COI* and **(B)** D-loop.

Germplasm Protection Mechanism

Historical records showed that the Qinghai Lake was originally connected to the ancient Yellow River, but after the strong uplift of the Qinghai-Tibet Plateau, the water could not be drained to the Yellow River, so the Qinghai Lake was separated from the Yellow River and became a closed inland lake (Chen et al., 1964; Li et al., 1996). The Qinghai Lake provided opportunities for the lineage diversification of *G. przewalskii* (Zhao et al., 2006). Although the lake area has increased in recent years, this established geographic barrier combined with the continued decline in water level, increased mineralization of the water, and shrinking of the water column in early years has led to a reduction in the size of spawning areas, a decrease in the numbers of the spawning population, a single genetic structure, low genetic diversity, and even the risk of inbreeding (Zhao et al., 2005; Liu et al., 2019). A large number of seedlings are put into the Qinghai Lake every year to replenish the stock and expand the size of the population. Fishery stocking has restored the naked carp resources in the Qinghai Lake to some extent. However, if the parents from the same place or same batch are used for breeding for many years, negative genetic effects such as genetic drift are likely to occur (Araki and Schmid, 2010).

There is a definite need for preventing inbreeding and maintaining population heritability. A reasonable approach to tackle this issue could be to rationally introduce multi-water parents and regularly expand and replace the parental population during artificial breeding and release. More specifically, parents can be Qinghai Lake naked carp from different geographic regions, such as the Lake Crooke and the seven geographical regions mentioned in this study. It also may be different subspecies of *G. przewalskii*, such as *G. przewalskii ganzilronensis*, and even other *Gymnocypris* fishes, such as *G. herzensten*. On the other hand, translocation is a usual measure for the effective conservation of imperiled species (Schäfer et al., 2020).

Studies on the adaptation of introduced *G. przewalskii* in other semi-saline lakes in the Qinghai Province are conducive to successfully protecting genetic diversity by creating various evolutionary units.

In conclusion, both markers indicated high haplotypic diversity, low nucleotide diversity, and low differentiation of *G. przewalskii*. With the launch of the "10-year fishing ban" and the sixth embargo fishing of the Qinghai Lake in 2020, it will play a considerable role in the conservation and restoration of the germplasm resources of *G. przewalskii*. Therefore, scientific management must be achieved in stock enhancement to avoid damaging the natural genetic structure of *G. przewalskii*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by the Regulations for the Administration of Affairs Concerning Experimental Animals, approved and authorized by the State Council of People's Republic of China.

AUTHOR CONTRIBUTIONS

D-aF was responsible for data scoring and analyses, and writing the manuscript. HL, CcM, MH, ZK, HfQ, DpX, LfT, and YdL helped in selecting the fish sample and data analysis during the manuscript preparation. All authors have read and approved the final manuscript.

FUNDING

This study was supported by the Fishery Resources and Environmental Investigation Special Qinghai Lake in Key Waters of Northwest China (2019–2021) and the National Infrastructure of Fishery Germplasm Resources (2021DKA3047003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.827654/full#supplementary-material

REFERENCES

Anneville, O., Lasne, E., Guillard, J., Eckmann, R., Stockwell, J. D., Gillet, C., et al. (2015). Impact of fishing and stocking practices on coregonid diversity. *Food Nutr. Sci.* 6, 1045–1055. doi: 10.4236/fns.2015.611108

Araki, H., and Schmid, C. (2010). Is hatchery stocking a help or harm?: evidence, limitations and future directions in ecological and genetic surveys. *Aquaculture* 308(Suppl.–S1), S2–S11.

Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Cao, Y.-B., Chen, X.-Q., Wang, S., Chen, X.-C., Wang, Y.-X., Chang, J. P., et al. (2009). Growth hormone and insulin-like growth factor of naked carp (*Gymnocypris przewalskii*) in Lake Qinghai: expression in different water environments. *Gen. Comp. Endocrinol.* 161, 400–406. doi: 10.1016/j.ygcen.2009.02.005

Chan, J., Li, W., Hu, X., Liu, Y., and Xu, Q. (2016). Genetic diversity and population structure analysis of Qinghai-Tibetan plateau schizothoracine fish (*Gymnocypris dobula*) based on mtDNA D-loop sequences. *Biochem. Syst. Ecol.* 69, 152–160. doi: 10.1016/j.bse.2016.09.004

Chen, D., Zhang, C., Lu, C., and Zhang, X. (2006). Polymorphism of D-loop sequence from mitochondrial genomes of different brood- stocks of *Gymnocypris przewalskii* (Kessler). *J. Fish. Sci. China* 13, 800–806.

Chen, K., Huang, D., and Liang, D. (1964). The formation and development of Qinghai Lake. *Acta Geogr. Sin.* 30, 214–233.

Cross, T. F. (2015). Genetic implications of translocation and stocking of fish species, with particular reference to Western Australia. *Aquac. Res.* 31, 83–94. doi: 10.1046/j.1365-2109.2000.00439.x

Donaldson, K. A., and Wilson, R. R. (1999). Amphi-Panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Mol. Phylogenet. Evol.* 13, 208–213. doi: 10.1006/mpev.1999.0625

Duan, Z., Zhao, K., Peng, Z., Li, J., Diogo, R., Zhao, X., et al. (2009). Comparative phylogeography of the Yellow River schizothoracine fishes (Cyprinidae): vicariance, expansion, and recent coalescence in response to the Quaternary environmental upheaval in the Tibetan Plateau. *Mol. Phylogenet. Evol.* 53, 1025–1031. doi: 10.1016/j.ympev.2009.03.034

Feng, C., Yu, L., Fu, S., Wei, P., and Zhou, W. (2021). Cultivation measures of "spring sliced" fish species in naked carp pond in Qinghai Lake. *Sci. Fish Farm.* 3, 11–12.

Feng, C., Yu, L., Luo, Y., Wu, H., and Yan, L. (2019). Water quality conditioning technique in fingerling of Qinghai Lake naked carp. *Sci. Fish Farm.* 5, 7–9.

Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709. doi: 10.1093/genetics/133.3.693

Gang, S., Yanhua, Q., Zuohua, Y., Shouhsien, L., Naifa, L., and Fumin, L. (2009). Phylogeography of the *Alcippe morrisonia* (Aves: Timaliidae): long population history beyond late Pleistocene glaciations. *BMC Evol. Biol.* 9:143. doi: 10.1186/1471-2148-9-143

Gatt, M. H., Ferguson, M. M., and Liskauskas, A. P. (2000). Comparison of control region sequencing and fragment RFLP analysis for resolving mitochondrial DNA variation and phylogenetic relationships among Great Lakes Walleyes. *Trans. Am. Fish. Soc.* 129, 1288–1299. doi: 10.1577/1548-8659(2000)129<1288:cocrsa>2.0.co;2

Gong, S., Wu, J., Gao, Y., Fong, J. J., Parham, J. F., and Shi, H. (2020). Integrating and updating wildlife conservation in China. *Curr. Biol.* 30, R915–R919. doi: 10.1016/j.cub.2020.06.080

Grant, W., and Bowen, B. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J. Heredity* 89, 415–426. doi: 10.1093/jhered/89.5.415

He, W. (2008). The cause of exhaustion of *Gymnocypris przewalskii* resources and the way of recovery. *Sci. Fish Farm.* 2, 5–85.

Jiang, H., Wu, J., Wu, Y., Qi, H., Zhou, W., Fu, S., et al. (2021). Effect of alizarin complexone on otolith marking of juvenile Qinghai-lake naked carp. *Sci. Fish Farm.* 3, 72–74.

Jiang, P., Shi, J., Zhang, Y., Qi, H., and Sun, X. (2009). Microsatellite variation analysis ofgenetic diversity in six wild populations of naked common carp. *Acta Ecol. Sin.* 29, 939–945.

Julio, R., Albert, F. M., Carlos, S. D. J., Sara, G. R., Pablo, L., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Laurent, E., and Heidi, L. (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

Li, J., Fang, X., Ma, H., Zhu, J., Pan, B., and Chen, H. (1996). Geomorphological and environmental evolution in the upper reaches of the Yellow River during the late Cenozoic. *Sci. China* 39, 380–390.

Liu, J. (2004). A quantitative analysis on threat and priority of conservation order of the endemic fishes in upper reaches of the Yangtze River. *China Environ. Sci.* 24, 395–399.

Liu, Q., Sha, Z., Wang, Q., Li, Q., and Ma, Y. (2019). "The variation characteristics and climate response of Qinghai Lake area in the past 32 years," in *Proceedings of the 2019 World Congress on Computational Intelligence, Engineering and Information Technology (WCEIT 2019)*, San Francisco, CA, 445–452.

Luo, L. (2015). *The Protection of Qinghai Lake Naked Carp Must Also be Added With a Heavy Fist*. Qinghai: Qinghai University for Nationalities.

Millar, C. I., Libby, W. J., Falk, D. A., and Holsinger, K. E. (1991). "Strategies for conserving clinal, ecotypic, and disjunct population diversiv in widespread species," in *Genetics and Conservation of Rare Plants*, eds D. A. Falk and K. E. Holsinger (Oxford: Oxford University Press), 149–170.

Min, Q., Zhang, X., Zhou, Q., Li, J., Zeng, S., Chen, F., et al. (2020). Cloning and preliminary function analysis of foxk1 gene in *Gymnocypris przewalskii*. *Guizhou Agric. Sci.* 48, 73–76.

Ministry of Agriculture and Rural Affairs of the People's Republic of China (2003). *Naked carp of Qinghai lake (SC 1063-2003)*. Beijing: Ministry of Agriculture and Rural Affairs of the People's Republic of China.

Miya, M., and Nishida, M. (2000). Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Mol. Phylogenet. Evol.* 17, 437–455. doi: 10.1006/mpev.2000.0839

Perez-Enriquez, R., Takagi, M., and Taniguchi, N. (1999). Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. *Aquaculture* 173, 413–423. doi: 10.1016/s0044-8486(98)00469-4

Qi, D. (2002). *Study on Genetic Diversity of Qinghai-Lake Naked Carps*. Hangzhou: Zhejiang University.

Rogers, A. R., and 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

Rômulo, P., Josiane, R., and Evoy, Z.-F. (2018). Genetic variation of the endangered neotropical catfish steindachneridion scriptum (Siluriformes: Pimelodidae). *Front. Genet.* 9:48. doi: 10.3389/fgene.2018.00048

Sari, A., Duran, M., Sen, A., and Bardakci, F. (2015). Investigation of Chironomidae (Diptera) relationships using mitochondrial COI gene. *Biochem. Syst. Ecol.* 59, 229–238. doi: 10.1006/mpev.2001.0898

Schäfer, D., Vincent, H., Fischer, M., and Kempel, A. (2020). The importance of genetic diversity for the translocation of eight threatened plant species into the wild. *Glob. Ecol. Conserv.* 24:e01240. doi: 10.1016/j.gecco.2020.e01240

Shi, J. (2008). Status of research on naked carp in Qinghai Lake and measures for resource conservation. *Qinghai Sci. Technol.* 05, 13–16.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. doi: 10.1093/genetics/123.3.585

Tong, C., Zhang, C., Shi, J., Qi, H., Zhang, R., Tang, Y., et al. (2015). Characterization of two paralogous myostatin genes and evidence for positive selection in Tibet fish: *Gymnocypris przewalskii*. *Gene* 565, 201–210. doi: 10.1016/j.gene.2015.04.009

Wang, L., Shi, X., Su, Y., Meng, Z., and Lin, H. (2013). Genetic divergence and historical demography in the endangered large yellow croaker revealed by mtDNA. *Biochem. Syst. Ecol.* 46, 137–144. doi: 10.1016/j.bse.2012.09.021

Wang, S., and Xie, Y. (2004). *China Species Red List*. Beijing: Higher Education Press.

Wang, Y., Jin, Z., Zhou, L., Li, F., Zhang, F., Chen, L., et al. (2014). Stratigraphy and otolith microchemistry of the naked carp *Gymnocypris przewalskii* (Kessler) and their indication for water level of Lake Qinghai during the Ming Dynasty of China. *Sci. China* 57, 2512–2521. doi: 10.1007/s11430-014-4836-1

Wang, Z., Shi, J., Wei, F., Qi, D., Shi, Y., Zhang, Y., et al. (2015). AFLP genmetic diversity of scaleless carp (*Gymnocypris przewalskii*) in different freshwater tributaries of Qinghai Lake. *Genom. Appl. Biol.* 34, 82–89.

Wu, X. (2017). *The Morphologic Feature and mtDNA Phylogeny of Rhinogobius SP*. Hefei: Anhui Agricultural University.

Wu, X., Cao, W., Yi, B., Yang, G., and Luo, Y. (1964). *Records of Cyprinidae fishes in China (Volume I and II)*. Hefei: Anhui Agricultural University, 181–182.

Wu, Y., Jiang, H., and Fu, S. (2020). Commercial fish culture technology of naked carp in Qinghai Lake in industrialized circulating waterwheel workshop. *Sci. Fish Farm.* 04, 82–83.

Xu, S., Li, T., Li, J., Zhu, S., and Zhao, K. (2003). Mitochondrial DNA diversity in naked carps in Qinghai Lake. *Heilongjiang J. Anim. Sci. Vet. Med.* 01, 11–12.

Yan, T., Wang, X., Li, S., He, Z., Luo, J., Zhang, Q., et al. (2020). Genetic analysis of wild *Ancherythroculter nigrocauda* in tributaries and the main stream of the upper Yangtze River basin of China. *Mitochondrial DNA A* 31, 17–24. doi: 10.1080/24701394.2019.1698558

Yang, T., Meng, W., Guo, Y., and Ma, Y. (2013). Morphological and COI sequences analysis of two schizothoracinae fishes in Xinjiang. *Chinese J. Fish.* 26, 10–14. doi: 10.3724/SP.J.1005.2014.1013

Yang, T., Meng, W., Zhang, R., Guo, Y., Hai, S., and Xie, C. (2011). Comparative analysis of mtDNA gene between two rare schizothoracine fishes. *Chinese J. Zool.* 46, 47–54.

Zhang, C., Chen, D., Shi, J., Qi, H., and Lu, C. (2005). Genetic diversity analysis of different populations of *Gymnocypris przewalskii* by RAPD. *J. Fish. China* 29, 307–312.

Zhang, C., and Zhang, Y. (1963). New species of fishes from Qinghai II. *Acta Zool. Sin.* 04, 146–149.

Zhang, Q., Sun, C., Zhu, Y., Xu, N., and Liu, H. (2019). Genetic diversity and structure of the round-tailed paradise fish (*Macropodus ocellatus*): implications for population management. *Glob. Ecol. Conserv.* 21:e00876. doi: 10.1016/j.gecco.2019.e00876

Zhang, R., Li, G., Tang, Y., Zhang, C., and Zhao, K. (2013). Genetic diversity and differentiation of mitochondrial DNA D-loop region in *Gymnocypris przewalskii*. *Chinese Agric. Sci. Bull.* 29, 71–76.

Zhang, X., Gao, X., Wang, J., and Cao, W. (2015). Extinction risk and conservation priority analyses for 64 endemic fishes in the Upper Yangtze River, China. *Environ. Biol. Fishes* 98, 261–272. doi: 10.1007/s10641-014-0257-4

Zhao, K., He, S., Peng, Z., and Li, J. (2006). Mitochondrial diversity and population structure of *Gymnocypris przewalskii* (Teleostei:Cyprinidae). *J. Qinghai Univ.* 24:5.

Zhao, K., Li, J., Yang, G., Duan, Z., He, S., and Chen, Y. (2005). Molecular phylogenetics of *Gymnocypris* (Teleostei: Cyprinidae) in Lake Qinghai and adjacent drainages. *Chinese Sci. Bull.* 50, 1325–1333. doi: 10.1360/982005-223

Zhao, K., Li, J., Zhang, Y., Luo, J., Li, T., Wu, H., et al. (2001). Mitochondrial DNA diversity in naked carps in Qinghai-lake (*Gymnocypris prezwalskii*). *Hereditas* 23, 445–448.

Zheng, B., Xu, Q., and Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quatern. Int.* 97, 93–101. doi: 10.1016/s1040-6182(02)00054-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fgene.2022.843848

Population Structure, Genetic Diversity, and Conservation Strategies of a Commercially Important Sleeper Fish, *Odontobutis potamophilus* (Gobiiformes: Odontobutidae) Based on Gene-Capture Data

Yun Hu[1,2], *Hongjie Li*[1,2], *Jianhong Xia*[3]* and *Chenhong Li*[1,2]*

[1]Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai, China, [2]Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, Shanghai, China, [3]Shanghai Natural History Museum, Branch of the Shanghai Science and Technology Museum, Shanghai, China

OPEN ACCESS

Edited by:
Narongrit Muangmai, Kasetsart University, Thailand

Reviewed by:
Sofia Priyadarsani Das, Central Institute of Freshwater Aquaculture (ICAR), India
Agus Nuryanto, Jenderal Soedirman University, Indonesia

***Correspondence:**
Jianhong Xia
snhmichth@hotmail.com
Chenhong Li
chli@shou.edu.cn

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 27 December 2021
Accepted: 21 April 2022
Published: 24 May 2022

Citation:
Hu Y, Li H, Xia J and Li C (2022) Population Structure, Genetic Diversity, and Conservation Strategies of a Commercially Important Sleeper Fish, Odontobutis potamophilus (Gobiiformes: Odontobutidae) Based on Gene-Capture Data. Front. Genet. 13:843848. doi: 10.3389/fgene.2022.843848

Odontobutis potamophilus is a popular food fish in China, distributed mainly in the middle and lower reaches of the Yangtze River, where it is a famous delicacy and a newly focused species for aquaculture. The wild populations of *O. potamophilus* are facing the problem of overfishing and habitat degradation. Therefore, it is very necessary to investigate and protect the wild populations of *O. potamophilus*. In this study, 72 fish were sampled from 18 different sites over its distribution range. Nuclear sequence data of 4,267 loci were collected using a gene-capture method. Phylogenetic reconstruction revealed that there were three major clades: Oujiang clade (OJ), Qiantang and lower Yangtze clade (QY), and middle Yangtze clade (MY). The discriminant analysis of principal components (DAPC) and a STRUCTURE analysis confirmed that there are three major groups within *O. potamophilus*. A fastsimcoal2 analysis corroborated the population history and suggested that there was discernible gene flow among these three groups, especially between QY and MY. Estimated pairwise F_{ST} suggested that Linhai (LH) and Shexian (SX) populations were the most divergent pair (F_{ST} = 0.7077). Taking the nucleotide diversity, population divergence, and admixture status altogether into consideration, we recommend that the LH, Gaoyou (GY) and Chaohu (CH) populations could be protected as the preferred resource for breeding projects. According to the results of genetic analyses, all populations of *O. potamophilus* should be protected due to low genetic diversity.

Keywords: *Odontobutis potamophilus*, population structure, conservation, genetic diversity, gene capture, genetic resource

INTRODUCTION

Genus *Odontobutis* contains a group of freshwater sleepers endemic to China, Japan, and the Korea Peninsula (Wu, 2008). There are eight species within this genus, *O. haifengensis*, *O. hikimius*, *O. interrupta*, *O. obscura*, *O. platycephala*, *O. potamophilus*, *O. sinensis*, and *O. yaluensis* (Froese and Daniel, 2014). In China, the two most interested species are *O. potamophilus* and *O. sinensis* because of their high economic value, so most of the research was focused on the reproduction and breeding of these two species (Liu et al., 2008; Zhao et al., 2009). Li et al. (2017) compared the average weight and absolute fecundity between *O. potamophilus* and *O. sinensis* found from Jiangxi Province of China, and found that *O. potamophilus* was more suitable for artificial propagation and aquaculture development than *O. sinensis*.

Odontobutis potamophilus is an economically important fish endemic to China. It is mainly distributed in the middle and lower reaches of the Yangtze River, as well as the Huaihe River, the Qiantang River, and the Oujiang River basins (Wu, 2008). *Odontobutis potamophilus* is small carnivorous fish that lives at the bottom of lakes, rivers, and ditches. It prefers to inhabit shallow waters where sand, weeds, and gravel are mixed (Wu, 2008). This fish has high meat-to-body weight ratio (64.45%) (Hu et al., 2012), fine meat quality, and a delicious flavor, but its natural yield is low. Therefore, *O. potamophilus* has become an attractive species for aquaculture in recent years (Wang et al., 2016).

With the increasing market demand, the breeding population of *O. potamophilus* has been in short supply (Zhang et al., 2019). In addition to being caught directly for the market, wild *O. potamophilus* has also been caught for the cultivation of cultured parent fish (Zhang et al., 2012). Driven by the rising price (USD 25 per Kg); the fishing enthusiasm of fishermen is unprecedented (Zhang et al., 2012; Zhang et al., 2019). With the increasing pollution of natural waters, the living environment of *O. potamophilus* was deteriorating and the spawning environment was destroyed repeatedly (Zhu et al., 2014; Wang et al., 2018). As a result of many factors, the natural resources of *O. potamophilus* were increasingly exhausted and the populations declined seriously (Zhao and Sheng, 2011). In the 2017 IUCN Red List, *O. potamophilus* was rated as an information deficient (Data Deficient, DD) species (Osipova et al., 2017). Analyzing the population structure of *O. potamophilus* is the first step to conserver its wild genetic diversity. Wang et al. (2016) hybridized three different geographical populations of *O. potamophilus* collected from Jiande, Zhejiang Province, Dangtu, Anhui Province, and Taihu, Jiangsu Province. They found that the growth rate of F1 hybrids between different populations was all higher than self-propagated within the same population. Therefore, investigating the genetic diversity of different populations also can provide more favorable germplasm resources for aquaculture. There is an urgent need for the estimation of genetic resources of *O. potamophilus*.

At present, most genetic studies on *O. potamophilus* were based on mitochondrial loci (Xu et al., 2015; Zhang et al., 2019), but see Zhu et al. (2014) for developing microsatellite markers for this species and population study covering its partial distribution. They discribed the genetic diversity of *O. potamophilus* in five populaitons, and suggested that *O. potamophilus* in those areas can be used for breeding and conservation programs (Zhu et al., 2014). Hou et al. (2014) show that genetic diversity was low (overall Pi = 0.04028) in *O. potamophilus*, and there was a high degree of differentiation (average F_{ST} = 0.92911) among populations, absenting gene flow between populations. However, in the studies of Hou et al. (2014) and Zhu et al. (2014), not the whole distribution range of the species was covered. Some populations of *O. potamophilus* with low genetic diversity were found (Hou et al., 2014; Zhu et al., 2014). One of the reasons for low genetic diversity could be inbreeding. Low genetic diversity will cause serious ecological consequences, such as reducing the ability of the population to survive in a stressful or changing environment, and eventually extinction (Hughes et al., 2008). Therefore, it is necessary to protect *O. potamophilus* with low genetic diversity. Most populations of the *O. potamophilus* from different geographic regions have not expanded (Hou et al., 2014). It is important to use multiple nuclear gene loci to evaluate the genetic resource of *O. potamophilus* in its whole distribution range.

Various methods have been applied for collecting thousands of nuclear loci for population studies, such as genotyping-by-sequencing (GBS), restriction site associated DNA sequencing (RAD Sequencing) (Baird et al., 2008; Elshire et al., 2011). Another strategy of reduced representative sequencing is through gene capture. Jiang et al. (2019) developed a suite of 4.434 single-copy nuclear gene markers for ray-finned fishes. Those markers can be used to study both phylogenomics and population genomics of ray-finned fishes, and those have been successfully tested in *Odontobutis* species (Li et al., 2018; Jiang et al., 2019).

In this study, we collected 72 fish from 18 different sites of the Huaihe River, the Yangtze River, the Qiantang River, and the Oujiang River basins. We applied a target-gene enrichment approach (Li et al., 2013) and the next generation sequencing method to obtain a large amount of nuclear gene data, and carried out on estimation of the genetic diversity and population structure of *O. potamophilus*. The results of this research can provide a reference for the selective breeding programs of *O. potamophilus* with high genetic diversity or populations with low genetic diversity but unique in terms of genetic composition.

MATERIALS AND METHODS

Sampling and DNA Extraction

A total of 72 *O. potamophilus* were collected from Meicheng (MC), Cixi (CX), Hangzhou (HZ), Huzhou (HU), Baoying (BY), Gaoyou (GY), Jingjiang (JJ), Guangfu (GF), Jiangba (JB), Shanghai (SH), Meizhu (MZ), Xingfu (XF), Shuiyang (SY), Chaohu (CH), Zhongmiao (ZM), Luan (LA), Shexian (SX), and Linhai (LH) of the Huaihe River, the Yangtze River, the Qiantang River, and the Oujiang River basins (**Table 1**; **Figure 1**). Two samples of *O. yaluensis* were used as outgroup. Most samples were collected before 2014 when the large-scale artificial breeding

TABLE 1 | Sampling sites and date of collection.

Code	Sample ID	No. of samples	Water body	Collection site	Date
MZ	CL2	3	Lake Nanyi, Yangtze River	Meizhu,Langxi, Anhui, China	2008
XF	CL3	2	Lake Nanyi, Yangtze River	Xingfu, Langxi, Anhui	2009
HU	CL364	5	Lake Tai, Yangtze River	Huzhou, Zhejiang, China	2013/4/7
SY	CL378	3	Shuiyangjiang, Yangtze River	Shuiyang, Xuancheng, Anhui, China	2013/4/7
CH	CL382	6	Lake Caohu, Yangtze River	Chaohu, Anhui, China	2013/4/10
ZM	CL384	3	Lake Caohu, Yangtze River	Zhongmiao,Hefei, Anhui, China	2013/4/10
LA	CL399	2	Huaihe River	Liuan, Anhui, China	2013/4/10
BY	CL411	4	Lake Baoying, Yangtze River	Baoying, jiangsu, China	2013/4/10
GY	CL421	4	Lake Gaoyou, Yangtze River	Gaoyou, Jiangsu, China	2013/4/10
JJ	CL423	5	Yangtze River	Jingjiang, Jiangsu, China	2013/4/12
GF	CL478	6	Lake Tai, Yangtze River	Guangfu, Jiangsu, China	2013/6/5
JB	CL482	2	Lake Hongze, Huaihe River	Jiangsu, China	2013/4/19
SH	CL495	4	Yangtze River	Shanghai, China	2013/07/04
SX	CL541	5	Yangtze River	Shexian, Huangshan, Anhui, China	2014/04/13
LH	CL763	5	Oujiang River	Linhai, Zhejiang, China	2014/6/26
MC	CL1221	5	Qiantang River	Meicheng, Zhejiang, China	2016/8/22
CX	CL1981	4	Qiantang River	Cixi, Ningbo, Zhejiang, China	2017/10/10
HZ	CL1982	4	Qiantang River	Hangzhou, Zhejiang, China	2017/10/12
Out group	CL850	1	Aihe, The Yalu River	Liaoning, China	2016/06/14
	CL1275	1	The Yalu River	Kuandian, Dandong, China	2016/11/12

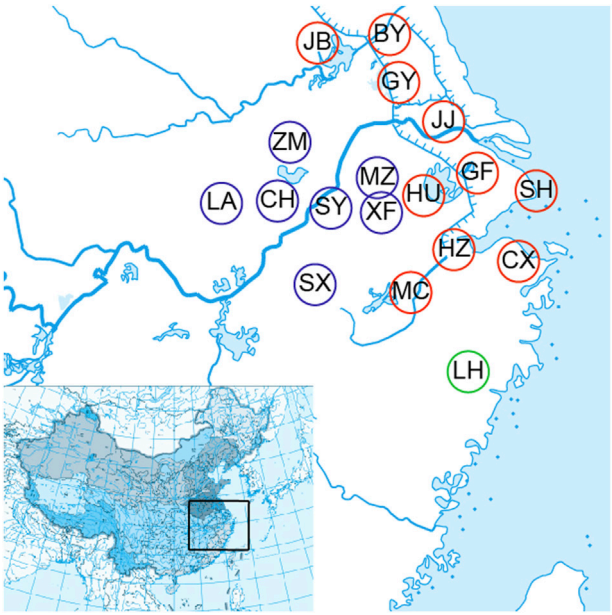


FIGURE 1 | Eighteen sample collecting sites of the *Odontobutis potamophilus* including Meicheng (MC), Cixi (Cx), Hangzhou (Hz), Huzhou (HU), Baoying (by), Gaoyou (Gy), Jingjiang (JJ), Guangfu (GF), Jiangba (JB), Shanghai (SH), Meizhu (MZ), Xingfu (XF), Chaohu (CH), Shuyang (SY), Zhongmiao (ZM), Liu (LA), Shexian (SX), and Linhai (LH).

and stocking started, so they are a good representation of the wild populations. The fin clips and muscle samples were taken and fixed with 75% ethanol before being brought back to the laboratory and stored at 4°C. All experimental procedures involving fish were approved by the Animal Ethics Committee of Shanghai Ocean University, China. Ezup Column Animal Genomic DNA Purification Kit (Sangon, Shanghai, China) was used to extract DNA from the tissue samples. The concentration of purified DNA was determined by a NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific. Wilmington, DE, United States) and visualized using agarose gel electrophoresis.

Library Construction, Gene Capture, and Sequencing

DNA was fragmented to ~250 bp using a Covaris M220 focused ultrasonicator (Woburn, Massachusetts, United States). One μl of sheared DNA was used in agarose gel electrophoreses to check the size of sheared DNA fragments. The Illumina sequencing libraries were constructed using a modified library preparation method (Meyer and Kircher, 2010; Li et al., 2013). In the ligation step of library preparation, each sample was labeled by adding inline indices to reduce the potential risk of cross-contamination between samples in the subsequent gene capture steps. Inline indices are short 6 bp nucleotide fragments that are connected to the ends of Illumina sequencing primers. The DNA libraries were mixed equimolarly for the subsequent gene capture experiments.

According to the 4,434 single-copy nucleotide coding sequence markers of bony fish developed by Jiang et al. (2019), the RNA baits were redesigned and the area with extreme high read depth in the preliminary experiment was masked. These biotin-labeled RNA baits were synthesized based on the sequence of a freshwater goby, *Rhinogobius giurinus*. The baits' length was 80 bp, covering the target sequences with 2X tiling. MYbaits RNA baits targeting the 4,434 loci were synthesized at Arbor Biosciences (Mycroarray, Ann Arbor, Michigan; cat#, 150901-Li-Goby).

Gene capture procedures were repeated twice followed Li et al. (2013). Briefly, sample libraries and baits were hybridized, and then the non-target fragments were washed off. The enriched libraries were amplified using IS4 and a P7 primer with an 8 bp DNA index following Meyer and Kircher (2010). After gene capture, the products were subjected to agarose gel electrophoreses for size selection, and products between 250 bp and 1,000 bp were cut and extracted. Gel extractions were sent to GENWIZ (Shanghai) for sequencing using an Illumina HiSeq platform.

Read Assembly

The raw Illumina data were distinguished by the 8 bp P7 barcodes and the 6 bp inline indices (Yuan et al., 2016). Trim_galore v0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, accessed on 10 July 2021) and cutadapt v1.2.1 (Martin, 2011) was used with default parameters to trim the adaptor sequence and reads with low-quality scores (Q < 20). The data assembly process was followed by Yuan et al. (2016). The paralogous sequences were filtered according to Yuan et al. (2016), and only loci captured for more than 30% of all individuals were kept for subsequent analyses. Clustal Omega was used to batch align the DNA sequence based on amino acid codes (Sievers et al., 2011).

SNP Calling and Population Genetic Metrics

After aligning the assembled results and removing low-quality sequences, pick_taxa.pl was used to construct consensus sequences as a reference for SNP calling, and SNPs from coding regions were targeted (Yuan et al., 2016). The cleaned reads were mapped to consensus sequences using bwa v0.7.16a-r1181 (Li and Durbin, 2009). Samtools were used to convert the same files into binary BAM files (Li et al., 2009). Picard (http://broadinstitute.github.io/picard) was used to remove PCR repeats. GATK-3.4.0 (McKenna et al., 2010) was used to genotype the SNP loci. Thirty individuals of SH, JB, GY, BY, JJ, HU, and GF were selected for the Hardy Weinberg equilibrium test using VCFtools (Danecek et al., 2011), because our preliminary analysis suggested those would be a panmictic population. Loci that did not conform to Hardy Weinberg equilibrium were identified and removed from VCF files generated for all individuals (72 samples). SNPs were screened with the parameters including MAF> 0.05, max-missing set as 0.90, and min-meANDP set as 6 using VCFtools. Only one SNP with the least missing data and the highest quality score was retained for each locus to meet the linkage balance requirements of subsequent analyses. The optimal SNP sites of the populations were selected, and the mean pairwise population differentiation (F_{ST}), nucleotide diversity (pi), and mean genetic distance (Phi_st) were calculated in the Stacks program (Catchen et al., 2013). According to the F_{ST} values between the two populations, the existence and degree of differentiation between the two populations were inferred (Wright, 1968).

Population Clustering and Genetic Structure

The captured genes were concatenated for phylogenetic tree inference. The phylogenetic tree was constructed using the maximum likelihood method to reveal the relationship of 72 *O. potamophilus* with two *O. yaluensis* as outgroups. The ML tree was reconstructed using RAxMLv8.0.0 under the GTRGAMMA model with 1000 bootstraps (Stamatakis, 2014). Figtree v1.4.2 was used to visualize the results (http://tree.bio.ed.ac.uk/software/Figuretree/).

STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to evaluate the population structure of the 72 individuals of *O. potamophilus*. The individuals were coded as 18 populations according to their geographical sampling sites. The STRUCTURE operating parameter was set to 20,000 for the length of the burn-in period, and 200,000 for the number of MCMC reps after burn-in. The analyses of k = 1–18 were repeated five times. The most likely K value was determined using Structure Harvester 0.6.93 (Earl, 2012).

The discriminant analysis of principal components (DAPC) was performed twice for each replication using R (Team, 2013) package adegenet version 2.1.1 (Jombart, 2008). The first time was used to evaluate the optimal number of PCs, and this value was used for the second analysis. The 72 individuals were divided into 18 groups by sampling sites for the cluster parameters. According to the number of PC lowest RMSE, the number of PCs chosen (≥1) was 10, and the number of distinct functions (≥1) was three. The scattering parameter was used for drawing.

History of Population Divergence

The results of cluster analysis showed that fish from the 18 sampling sites could be lumped into three major groups: The Oujiang clade (OJ), the Qiantang and lower Yangtze clade (QY), and the middle Yangtze clade (MY). We hypothesized four models for the evolutionary history of the three groups: model 1–3, either OJ, QY, or MY diverged first from the other two groups; model 4, the best model from the previous three but with migration allowed between different groups.

Different models were tested using fastsimcoal2 version 2.7 (Excoffier, 2011; Excoffier and Foll, 2011; Excoffier et al., 2021). Fastsimcoal2 can infer the population history by simulating historical events; it also can use the joint site frequency spectrum (JSFS) between populations as a summary statistics. Parameters include time, source, sink, migrants, new size, new growth rate, and migration matrix. The VCF file was converted into *.arp file using PGDSpider 2.1.1.5 (Lischer and Excoffier, 2012). The result of the Arp file was used as the input file for Alequin3.5 (Excoffier and Lischer, 2006) to calculate the site frequency spectrum (SFS). Based on the calculated SFS file (*.obs), different evolution scenarios were simulated in Fastsimcoal 2.7. Each model runs 100 independent point estimates, including 100 K coalescent simulation, and 30 ECM cycles ("-L 30"). The optimal JSFS observed by the model was compared with the Akaike information criterion (AIC). The conditional interval (CI) of the optimal model's point estimation was performed at 10 runs per bootstrapped SFS.

RESULTS

Next-Generation Sequencing Data

A total of 287,518,950 reads were generated, with an average of 3,993,319 reads per individual. A total of 223,207,564 reads were retained, after the adapter sequences and the low-quality

TABLE 2 | The nucleotide diversity (pi) of *Odontobutis potamophilus* of each population, including variant positions and all positions.

Group	Code	pi (variant positions)
QY	MC	0.1846
	CX	0.1289
	HZ	0.1391
	HU	0.1673
	BY	0.1374
	GY	0.1438
	JJ	0.1365
	GF	0.1461
	JB	0.1188
	SH	0.1562
MY	MZ	0.1119
	XF	0.1335
	CH	0.1330
	SY	0.1351
	ZM	0.1114
	LA	0.0487
	SX	0.0701
OJ	LH	0.0471

sequence (Q < 20) were trimmed, with an average of 3,100,106 reads per individual. After excluding PCR duplicates, 77.63% of the trimmed data were retained. An average of 2,635 loci were enriched in each sample, with the highest of 3,571 loci captured in CL1221_2, and the lowest of 202 loci in CL 1981_5. There were 3,750 loci on average captured for the two outgroup samples (**Supplementary Table S1**). After manual examination of all loci of *O. potamophilus* and *O. yaluensis*, 4,267 genes (96.23% of 4434 genes) were enriched in more than 22 (>30%) individuals and used for phylogenetic reconstruction.

Genetic Diversity and Differentiation

There were 12,988 SNPs called in total, with an average of 4.9 SNPs per locus. There were 6,045 out of 7,105 variants passing filters and HWE test in the 30 individuals examined. Those 1,060 sites that did not pass the HWE test were removed from the variants in the whole dataset (72 samples), keeping 11,932 out of the 12,988 SNPs. At least one SNP was called for 472 loci after selecting one SNP from each locus to avoid linkage disequilibrium. The average nucleotide diversity (pi) of variant positions was 0.1250 with the highest 0.1846 for MC, and the lowest 0.0471 found in LH (**Table 2**). Populations with large genetic difference ($F_{ST} > 0.25$) accounted for 60.13%, medium degree of genetic difference ($0.25 > F_{ST} > 0.15$) accounted for 16.34%, and existed genetic difference ($0.05 < F_{ST} < 0.15$) accounted for 23.53%. LH, ZM, LA, and SX were the most divergent populations from the others. The maximum genetic divergence was found between LH and SX ($F_{ST} = 0.7077$), and the lowest genetic differentiation was between GF and HU ($F_{ST} = 0.0492$) (**Table 3**).

Population Structure

The sampled sites can be divided into three groups (k = 3) (**Supplementary Figure S1**), including the Oujiang group (OJ, green), the Qiantang and lower Yangtze group (QY, red), and the middle Yangtze group (MY, blue) (**Supplementary Table S2**; **Figure 2**). Among the three groups, MC and CX populations had the

TABLE 3 | Pairwise differences (F_{ST}) among populations.

	MZ	XF	SY	LA	CH	ZM	SX	MC	CX	GF	HZ	GY	JJ	JB	BY	SH	HU
LH	0.6355	0.6377	0.5929	0.7501	0.5501	0.6591	0.7077	0.4404	0.5925	0.4839	0.5538	0.5652	0.5585	0.6907	0.5679	0.5380	0.4869
MZ		0.1309	0.1332	0.3709	0.1215	0.2005	0.2901	0.2583	0.3634	0.2770	0.3293	0.3418	0.3251	0.4451	0.3376	0.3198	0.2747
XF			0.1211	0.3585	0.1024	0.1802	0.2988	0.2441	0.3700	0.2729	0.3362	0.3458	0.3307	0.4562	0.3477	0.3163	0.2740
SY				0.2889	0.0926	0.1533	0.2062	0.2476	0.3387	0.2758	0.3222	0.3285	0.3163	0.4012	0.3243	0.3108	0.2725
LA					0.2080	0.3546	0.4373	0.2500	0.4330	0.2951	0.3822	0.3968	0.3730	0.5644	0.3989	0.3592	0.3002
CH						0.0805	0.1930	0.2476	0.3071	0.2657	0.2946	0.3064	0.2963	0.3513	0.3073	0.2926	0.2621
ZM							0.2884	0.2765	0.3856	0.3004	0.3607	0.3725	0.3539	0.4686	0.3672	0.3477	0.3067
SX								0.3452	0.4576	0.3677	0.4271	0.4361	0.4177	0.5311	0.4266	0.4173	0.3786
MC									0.1814	0.1579	0.1860	0.2033	0.1970	0.2075	0.2048	0.1809	0.1568
CX										0.1204	0.1758	0.1672	0.1615	0.2298	0.1704	0.1541	0.1283
GF											0.0829	0.0866	0.0829	0.1039	0.0887	0.0712	0.0492
HZ												0.1419	0.1305	0.1906	0.1382	0.1229	0.0827
GY													0.0821	0.1195	0.0748	0.1027	0.0840
JJ														0.1213	0.0823	0.1136	0.0880
JB															0.1253	0.1329	0.1047
BY																0.1133	0.0913
SH																	0.0679

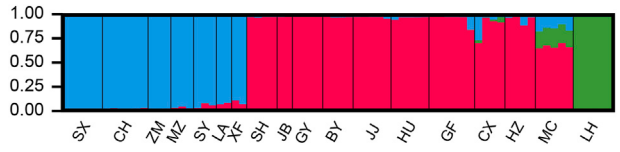


FIGURE 2 | The results of STUCTURE analysis of *Odontobutis potamophilus* populations based on 472 SNP loci. Red denotes the Qiantang and lower Yangtze group (QY), blue denotes the middle Yangtze group (MY), and green denotes the Oujiang clade (OJ). Each individual is represented by a single bar. The different color on a single bar indicates admixture origin of the individual.

most admixture individuals. These individuals were dominated by red loci, and green and blue loci only accounted for a small part. LH, GY, JB, and SX populations were highly independent, and they had almost no admixture with the other groups. Results of DAPC also showed that the populations of *O. potamophilus* can be divided into three groups (**Figure 3**). According to the results of the ML tree (**Figure 4**), the first clade split off was the Oujiang clade (OJ, green), and then the Qiantang and lower Yangtze clade (QY, red), and the middle Yangtze clade (MY, blue). This pattern cooperated with the results of STRUCTURE analysis and DAPC.

Historical Population Dynamics of *O. potamophilus*

The bestlhoods file generated using fastsimcoal2 analyses has two types of likelihoods, including MaxObsLhood and MaxEstLhood. MaxObsLhood is the maximum possible value for the expected SFS to match the observed SFS. MaxEstLhood is the maximum likelihood estimated according to model parameters. The higher the model fit with the population history, the smaller the difference between MaxObsLhood and MaxEstLhood. The MaxObsLhood values of all four models were −155.92 (**Table 4**). Model 1 had the lowest MaxEstLhood (−222.96) while model 2 had the largest (−228.8) (**Supplementary Figure S2**). Therefore, the most suitable model among the first three models was model 1. Based on the genetic structure of model 1, the parameters of gene flow were added to formulate model 4. Model 4 had smaller a MaxObsLhood (−217.9) than model 1 (**Table 4**), and Model 4 also had the lowest AIC value (1013.46), so we chose model 4 as the best model to explain the population history of *O. potamophilus* (**Table 4**). According to model 4, the common ancestor A_0 differentiated into OJ and A_1, and then A_1 differentiated into MY and QY clades (**Figure 5**). There was gene flow among all three clades, but MY and QY clades had higher gene flow. The gene flow between MY and QY clades was 100-fold larger than the gene flow between OJ and QY (or MY) (**Figure 6**).

DISCUSSIONS

Genetic Diversity and Candidate Populations for Conservation

Genetic diversity plays an important role in the survival and adaptability of species and it is also considered a basis for genetic breeding and conservation (Frankham, 2005; King and Lively, 2012). This study provides a framework for the conservation of *O. potamophilus*. The higher the genetic diversity of the populations, the stronger the adaptability to environmental changes. Therefore, populations with high genetic diversity are more suitable as founder population for a breeding program to be protected. The principle of selecting parents is to select populations with high genetic diversity and no admixed individuals in each group as ideal parents. The higher the genetic diversity, the richer the genetic information carried by the species, which is conducive to the subsequent breeding of excellent varieties. On the basis of high genetic diversity, we selected populations from each group without hybridized individuals from the other two groups, so as to avoid carrying the same genetic information between the two populations as parents. Thus, maintaining and improving the good quality of the progenies of *O. potamophilus*.

In the OJ group, LH was the only population sampled, with a nucleotide diversity of 0.0471. Mean pairwise population differentiation (F_{ST}) is an indicator to measure genetic differences between populations. The maximum value of F_{ST} was 0.7077 between LH and SX, indicating the greatest genetic difference between the two populations. In fact, the F_{ST} values between LH and all other populations were greater than 0.25,

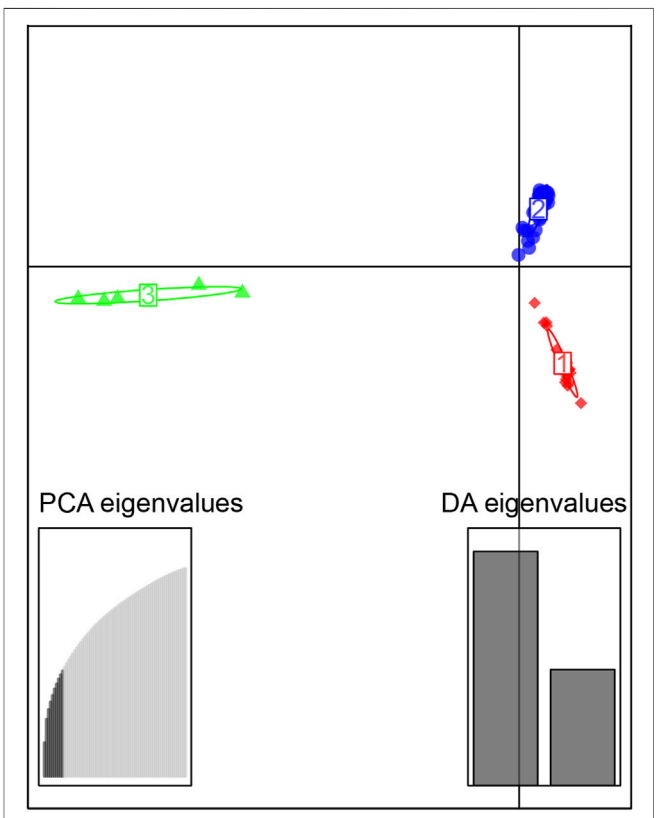


FIGURE 3 | Scatterplot generated by the discriminant analysis of principal components (DAPC). The 72 individuals were divided into 3 clusters, red, the Qiantang and lower Yangtze group (QY), blue, the middle Yangtze group (MY), and green, the Oujiang group (OJ).

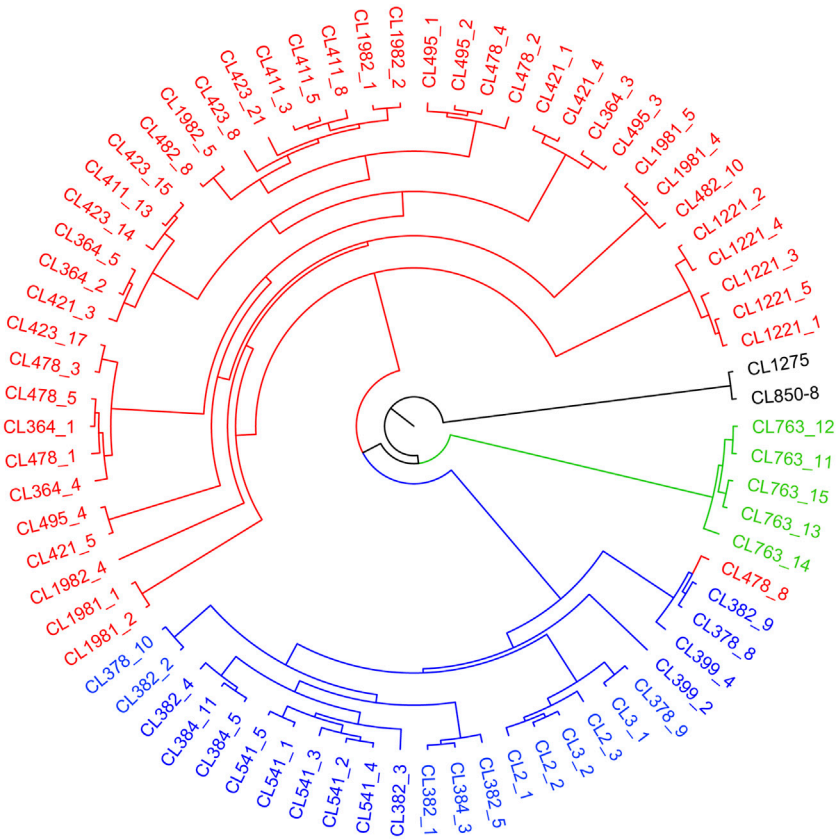


FIGURE 4 | A maximum likelihood tree based on captured genes using RAxML under GTRGAMMAR mode with 1,000 bootstrap replicates. The tree consists four parts, including red QY clade (the Qiantang and lower Yangtze clade), blue MY clade (the middle Yangtze clade), green OJ clade (the Oujiang clade), and black outgroup clade.

TABLE 4 | Comparing different models of population divergence history of *Odontobutis potamophilus* based on site frequency spectrum (SFS).

	MaxEstLhood	MaxObsLhood	AIC	Delta-likelihood
Modle1	−222.96	−155.92	1036.79	67.05
Modle2	−228.80	−155.92	1063.68	72.89
Modle3	−226.98	−155.92	1055.28	71.06
Modle4	−217.90	−155.92	1013.46	61.98

suggesting that LH is a unique genetic resource and breeding material. Therefore, LH should be protected for breeding programs, but more sites in the Oujiang River basin should be surveyed to examine the genetic diversity of *O. potamophilus* within the Oujiang River system. In the MY group, the population with the highest nucleotide diversity was SY (pi = 0.1351). However, many admixed individuals were observed in the SY population, so we should not use it for breeding programs. We suggest that the CH population, which had the second high nucleotide diversity (pi = 0.1330) but with no admixed individuals could be used for breeding projects. In the QY group, the population with the highest nucleotide diversity was MC (pi = 0.1846) but many admixed individuals were found in this population, so we suggest the GY (pi = 0.1438) population as the representative population due to the higher pi value and no admixed individuals. Taking the parameters of nucleotide diversity, F_{ST}, and admixture status all into consideration, we suggest that LH, GY, and CH populations should be protected. Those populations could be the preferred resource as founder populations for breeding projects, representing the OJ, QY, and MY groups, respectively.

Conservation is urgently needed in some populations, such as LA (Pi = 0.0487) and SX (Pi = 0.0701), due to their low genetic diversity compared to other populations. In these populations with low genetic diversity, the loss of individuals may lead to a decrease in the ability of the population to adapt to long-term environmental changes. Therefore, populations with low genetic diversity also should be protected, for example, by banning illegal fishing, to prevent anthropogenic destruction of the ecological environment, and creating a favorable habitat environment to maintain the wild populations of *O. potamophilus*.

Population Structure and History

Odontobutis potamophilus is distributed in the Huaihe River, the Yangtze River, the Qiantang River, and the Oujiang River basins; however, how many divergent groups of *O. potamophilus* in these

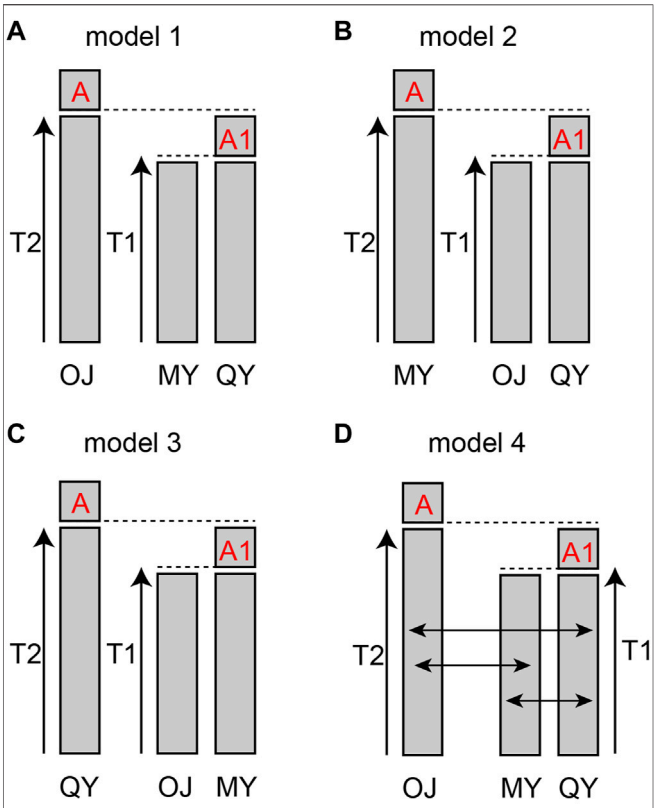


FIGURE 5 | Models of population history of *Odontobutis potamophilus*. **(A)** OJ split off first, and the other two groups differentiated later from a common ancestor A_0. **(B)** the first population diverged was MY. **(C)** the first population split off was QY. **(D)** the gene flow parameters added to model 1.

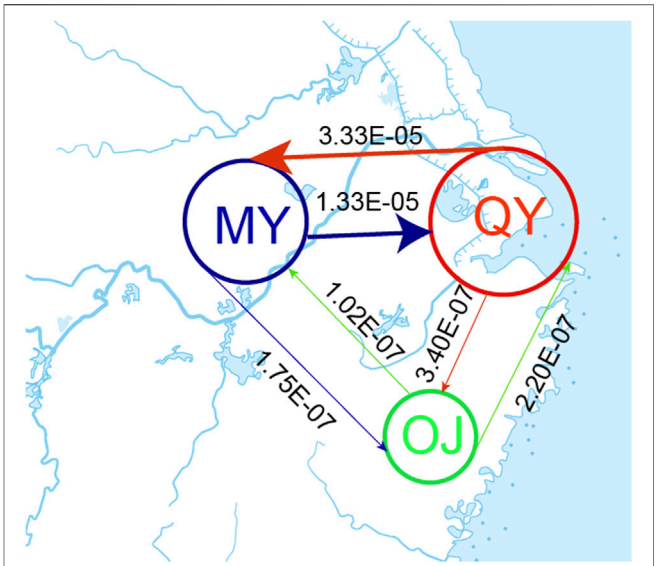


FIGURE 6 | Estimated gene flow between adjacent clades. The number on the line showed the strength of gene flow. The Oujiang clade (OJ), the Qiantang and lower Yangtze clade (QY), and the middle Yangtze clade (MY).

river systems have not been identified previously. The results of ML tree, DAPC, and STRUCTURE analyses unequivocally supported the division of *O. potamophilus* into three groups. The first differentiated group (OJ) was a clade of samples collected from the Oujiang River at Linhai, Zhejiang Province. Compared with other water systems, the Oujiang River is located in the southernmost part of the distribution range, so OJ population can be regarded as an independent population and it is crucial to protect the population of *O. potamophilus* in Oujiang. The other populations were divided into two groups, with Lake Tai as the boundary. The QY group is located to the east of Lake Tai, which is composed of populations from the Huaihe River, the Qiantang River, and lower Yangtze River. To the west of Lake Tai is MY group, which is composed of populations from the middle Yangtze River in Anhui Province. A similar pattern was also reported in previous research (Li, 2015). Most lakes in the middle and lower reaches of the Yangtze River dried up during the glacial periods due to the decline of sea level (Yang et al., 2000). The decline of sea level might lead to the isolation of the QY group in coastal shelters from the MY group. Lake Tai formed at about 3.7 Ka after the rise of sea level (Hong, 1991). During this period, Lake Tai might become a hybrid zone of QY and MY. The patterns of species differentiation and hybridization around Lake Tai also were found in other fishes, such as *Coilia nasus* (Cheng et al., 2019).

STRUCTURE results revealed that many populations of MY and QY had gene flow between groups. The geographical locations of these two groups are different sections of the Yangtze River. After the sea level rise, the previously isolated lakes were connected to the Yangtze. Therefore, individuals with admixed alleles of MY and QY were found in populations around Lake Tai, such as SY, XF, and GF. The individuals of LH (OJ group) had little gene flow with QY groups, probably because the Oujiang River has no direct connections with the Qiantang River or the Yangtze River. There were possibly few river capture events between the Oujiang River and the Qiantang River, therefore, a few individuals from MC and CX (the Qiantang River) showed admixture genotype with some alleles from the LH population (**Figure 2**).

Sample CL478-8 was collected from GF, which belongs to the QY group but clustered with individuals of the MY group in the phylogenetic tree (**Figure 4**). We speculate that CL478-8 was a hybrid between QY and MY. Moreover, the result of population history dynamics is consistent with the STRUCTURE results, that gene flow between MY and QY was the highest. We suggested that *O. potamophilus* should be divided into three management units, that is, the group to the east of Lake Tai, the group to the west of Lake Tai, and the Oujiang group.

Perspectives on Selective Breeding of *O. potamophilus*

If the volume of farmed *O. potamophilus* can meet consumer demand, the quantities of wild *O. potamophilus* caught can be reduced. Although artificial breeding of *O. potamophilus* has been

successful, there is yet no selected strain for aquaculture. Nonetheless, within-species crossing produced obvious heterosis in *O. potamophilus* (Wang et al., 2016), so there is high potential for selective breeding of *O. potamophilus*. A famous example of breeding based on within-species crossing and subsequent selection is the "GIFT tilapia". The successful breeding of GIFT tilapia indicates that the high genetic diversity of founder populations can improve the genetic gain of breeding (Ponzoni et al., 2005). We argue that a similar strategy can be applied to the selective breeding of *O. potamophilus*, that is, a founder population can be formulated by crossing LH, GY, and CH, the representative populations of the three groups to increase the background diversity of the founder population; and, then subsequent selection can be carried out on it. Meanwhile, the wild populations of the three groups should be conserved separately to keep the wild diversity of this fish.

CONCLUSION

In this study, we found that *O. potamophilus* could be divided into three major groups. Based on the nucleotide diversity, pairwise population differentiation, and admixture status of the populations analyzed, we recommended three candidates for the founder population and a strategy for selective breeding for *O. potamophilus*. All of the *O. potamophilus* populations should be considered for conservation to preserve their genetic diversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA792327.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Ethics Committee of Shanghai Ocean University, China.

AUTHOR CONTRIBUTIONS

JX, CL, and YH conceived the research plan. YH and HL did experiments and analyzed data. JX, CL, HL, and YH wrote the draft. All authors edited and approved the final version of the manuscript.

FUNDING

This work was supported by the Science and Technology Commission of Shanghai Municipality (19050501900) to CL and the Ministry of Science and Technology of the People's Republic of China (No. 2015FY110200) to JX.

ACKNOWLEDGMENTS

We are grateful to Zhizhi Liu, Jingyan Li, Ting Kuang, Qiang Li, Xin Zheng, and Peiyin Yang for helping with sample collection. We also thank the reviewers for their comments that helped in improving the quality of this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.843848/full#supplementary-material

REFERENCES

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* 3, e3376. doi:10.1371/journal.pone.0003376

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an Analysis Tool Set for Population Genomics. *Mol. Ecol.* 22, 3124–3140. doi:10.1111/mec.12354

Cheng, F., Wang, Q., Maisano Delser, P., and Li, C. (2019). Multiple Freshwater Invasions of the Tapertail Anchovy (Clupeiformes: Engraulidae) of the Yangtze River. *Ecol. Evol.* 9, 12202–12215. doi:10.1002/ece3.5708

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). Genomes Project AnalysisThe Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A Robust, Simple Genotyping-By-Sequencing (GBS) Approach for High Diversity Species. *PLoS One* 6, e19379. doi:10.1371/journal.pone.0019379

Excoffier, L. (2011). *A Continuous-Time Coalescent Simulator of Genomic Diversity under Arbitrarily Complex Evolutionary Scenarios*.

Excoffier, L., and Foll, M. (2011). Fastsimcoal: a Continuous-Time Coalescent Simulator of Genomic Diversity under Arbitrarily Complex Evolutionary Scenarios. *Bioinformatics* 27, 1332–1334. doi:10.1093/bioinformatics/btr124

Excoffier, L., and Lischer, H. (2006). *An Integrated Software Package for Population Genetics Data analysisComputational and Molecular Population Genetics Lab (CMPG)*. Switzerland: Institute of Zoology, University of Berne.

Excoffier, L., Marchi, N., and Sousa, V. C. (2021). *Fastsimcoal Ver 2.7 Fsc27*.

Frankham, R. (2005). Genetics and Extinction. *Biol. Conserv.* 126, 131–140. doi:10.1016/j.biocon.2005.05.002

Froese, R., and Pauly, D. (Editors) (202). *FishBase*. World Wide Web Electronic Publication. www.fishbase.org version (02/2022).

Hong, X. (1991). Origin and Evolution of the Taihu Lake [J]. *Mar. Geol. Quat. Geol.* 4.

Hou, X., Zhu, F., Yin, S., Zhang, L., Hu, Y., Wang, Y., et al. (2014). Genetic Diversity ofOdontobutis Potamophilafrom Different Geographic Populations Inferred from mtDNA Control Region. *Mitochondrial DNA* 25, 400–406. doi:10.3109/19401736.2013.803084

Hu, Y., Zhu, X., Yin, S., and Zhu, F. (2012). Analysis of Flesh Content and Muscle Nutritional Components of Odontobutis Potamophila. *Jiangsu Agric. Sci.* 40, 290–293.

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., and Vellend, M. (2008). Ecological Consequences of Genetic Diversity. *Ecol. Lett.* 11, 609–623. doi:10.1111/j.1461-0248.2008.01179.x

Jiang, J., Yuan, H., Zheng, X., Wang, Q., Kuang, T., Li, J., et al. (2019). Gene Markers for Exon Capture and Phylogenomics in Ray-finned Fishes. *Ecol. Evol.* 9, 3973–3983. doi:10.1002/ece3.5026

Jombart, T. (2008). Adegenet: a R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129

King, K. C., and Lively, C. M. (2012). Does Genetic Diversity Limit Disease Spread in Natural Host Populations? *Heredity* 109, 199–203. doi:10.1038/hdy.2012.33

Li, C., Hofreiter, M., Straube, N., Corrigan, S., and Naylor, G. J. P. (2013). Capturing Protein-Coding Genes across Highly Divergent Species. *Biotechniques* 54, 321–326. doi:10.2144/000114039

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, H., He, Y., Jiang, J., Liu, Z., and Li, C. (2018). Molecular Systematics and Phylogenetic Analysis of the Asian Endemic Freshwater Sleepers (Gobiiformes: Odontobutidae). *Mol. Phylogenetics Evol.* 121, 1–11. doi:10.1016/j.ympev.2017.12.026

Li, Y., Qiu, W., Wang, H., and Zhu, S. (2017). Morphylogical Characters and Reproductional Study on Odontobutis Species from Jiangxi Province. *Jiangxi Fish. Sci. Technol.*, 7–9.

Lischer, H. E. L., and Excoffier, L. (2012). PGDSpider: an Automated Data Conversion Tool for Connecting Population Genetics and Genomics Programs. *Bioinformatics* 28, 298–299. doi:10.1093/bioinformatics/btr642

Liu, M., Hu, X.-C., Han, Q., and Luo, Y. (2008). Variation of the Proteinic Enzymes Activities during the Embryonic and Larval Development of Odontobutis Potamophila. *Freshw. Fish.* 38, 39–41.

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* 17, 10–12. doi:10.14806/ej.17.1.200

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Meyer, M., and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* 2010, pdb. prot5448. doi:10.1101/pdb.prot5448

Osipova, E., Shadie, P., Zwahlen, C., Osti, M., Shi, Y., Kormos, C., et al. (2017). *IUCN World Heritage Outlook 2: A Conservation Assessment of All Natural World Heritage Sites*, 92. Gland, Switzerland: IUCN, Gland.

Ponzoni, R. W., Hamzah, A., Tan, S., and Kamaruzzaman, N. (2005). Genetic Parameters and Response to Selection for Live Weight in the GIFT Strain of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 247, 203–210. doi:10.1016/j.aquaculture.2005.02.020

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75

Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Team, R. C. (2013). *R: A Language and Environment for Statistical Computing*.

Wang, P., Ding, Y., Yu, X., Jia, X., Zhang, G., Yin, S., et al. (2016). Growth Trait and Genetic Diversity in the F1 Progeny of Dark Sleeper Odontobutis Potamophila from Three Different Populations. *Fish. Sci.* 35, 528–534.

Wang, P., Zhao, C., Chen, S., Yin, S., Wang, Q., Li, Z., et al. (2018). Parentage Identification of Odontobutis Potamophlia Based on Microsatellite DNA Markers. *J. Genet.* 97, 563–568. doi:10.1007/s12041-018-0933-9

Wright, S. (1968). "Evolution and the Genetics of Populations," in *Genetic and Biometric Foundations* (Chicago, IL: Univ. of Chicago Press), 1.

Wu, H., Zhong, J., Chen, I. S., Chong, D., Ni, Y., Shen, G., et al. (2008). *Fauna Sinica: Ostichthyes. Perciformes. Gobioidei. V*. Beijing, China: Science Press.

Xu, Y., Zhong, L., Li, X., Zhu, X., Wang, C., and Shi, Y. (2015). Genetic Variation Analysis of Odontobutis Potamophila from Five Lakes Based on Mitochondrial DNA Cyt B. *J. Lake Sci.* 27, 693–699.

Yang, D., Li, X., and Zhang, Z. (2000). Lake Evolution along Middle-Lower Reaches of the Yangtze River. *J. Lake Sci.* 12, 232–238.

Yuan, H., Jiang, J., Jiménez, F. A., Hoberg, E. P., Cook, J. A., Galbreath, K. E., et al. (2016). Target Gene Enrichment in the Cyclophyllidean Cestodes, the Most Diverse Group of Tapeworms. *Mol. Ecol. Resour.* 16, 1095–1106. doi:10.1111/1755-0998.12532

Zhang, G., Shi, Y., Zhang, H., Lu, G., and Qiao, Y. (2012). The Techniques of Raising Odontobutis Potamophila 1. Biology and Marketing Potentials. *Fish. Sci. Technol. Inf.* 39, 123–127.

Zhang, Y., Ling, J., Duan, G., Hu, Y., Pan, T., Zhou, H., et al. (2019). Genetic Diversity and Population Structure Analyses of Odontobutis Potamophila from Different Geographic Populations in Anhui Province. *J. Anhui Agric. Univ.* 46, 383–388.

Zhao, J., and Sheng, J. (2011). Experiments on Artificial Breeding and Seed Rearing of Odontobutis Potamophila. *Fish. Sci. Technol. Inf.* 38, 118–120.

Zhao, X.-Q., Chen, L.-Q., Gu, Z.-M., Li, E.-C., and Yao, Q. (2009). Effects of Starvation on Ovarian Development in Female Odontobutis Potamophila during Over-wintering Period. *J. Fish. China* 33, 70–77.

Zhu, F., Zhang, L.-J., Yin, S.-W., Zhang, H.-W., Hou, X.-Y., Hu, Y.-L., et al. (2014). Genetic Diversity and Variation in Wild Populations of Dark Sleeper (Odontobutis Potamophila) in China Inferred with Microsatellite Markers. *Biochem. Syst. Ecol.* 57, 40–47. doi:10.1016/j.bse.2014.07.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Ecology and Evolution

BRIEF RESEARCH REPORT
published: 07 June 2022
doi: 10.3389/fevo.2022.865089

Reanalysis on Phylogeographic Pattern of Sharpbelly *Hemiculter leucisculus* (Cyprinidae: Cultrinae) in China: A Review and the Implications for Conservation

Qianhong Gu[1,2†], Hui Zhong[1†], Yaxian Sun[1†], Hui Yuan[1†], Shengnan Li[1], Zhongyuan Shen[1] and Ming Wen[1]*

[1] State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Sciences, Hunan Normal University, Changsha, China, [2] Engineering Technology Research Center of Henan Province for Aquatic Animal Cultivation, College of Fisheries, Henan Normal University, Xinxiang, China

OPEN ACCESS

Edited by:
Shaokui Yi,
Huzhou University, China

Reviewed by:
Min Li,
Chinese Academy of Fishery Sciences (CAFS), China
Zhong Jia,
Yangtze River Fisheries Research Institute, China

***Correspondence:**
Ming Wen
ming.wen@hunnu.edu.cn;
gqh@hunnu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: 04 February 2022
Accepted: 25 April 2022
Published: 07 June 2022

Citation:
Gu Q, Zhong H, Sun Y, Yuan H, Li S, Shen Z and Wen M (2022) Reanalysis on Phylogeographic Pattern of Sharpbelly Hemiculter leucisculus (Cyprinidae: Cultrinae) in China: A Review and the Implications for Conservation. Front. Ecol. Evol. 10:865089. doi: 10.3389/fevo.2022.865089

Hemiculter leucisclus, as a widely distributed freshwater fish in China, provides an interesting model to explore the impact of drainage evolution and geologic history in the Pleistocene on diversification patterns. We collected the mitochondrial cytochrome b (cytb) gene and the recombination activating gene 2 (RAG2) from 1,070 individuals from 59 sampling locations. Phylogenetic and population genetic approaches were used to describe the phylogeographic pattern and to test how the geological and climatic factors on diversification. The results suggested that there existed four sublineages of the *H. leucisclus* across six river systems, among which two sublineages, showing strongly indigenous characteristics, are constrained to particular geographical regions in China. The molecular data and ancestral states demonstrated that the *H. leucisclus* possibly originated from the Pearl River basin during the later Pliocene. The phylogeographic pattern in *H. leucisclus* appears to have been driven by palaeoenvironmental perturbations rather than anthropogenic translocations. The geographically constrained sublineages A in the middle and lower Pearl River basin and sublineage B in the upper Yangtze River basin deserves special protection.

Keywords: phylogeographic pattern, phylogenetic tree, divergence time, Pliocene and Pleistocene, sharpbelly

INTRODUCTION

Landscape features and geological history are hypothesized to strongly affect the biogeographical processes (Manel et al., 2003), and this can be especially true in freshwater systems (Leclerc et al., 2008). The river system evolution can shape and influence the distributions and phylogeographical pattern of freshwater-limited lineages among drainages (Schultheiß et al., 2014). The population structure and diversification of freshwater fishes are primarily influenced by the dynamics of river morphology, affecting connections between river basins and lakes. The separation of water bodies often leads to genetic divergence as a result of genetic drift and reproductive isolation (Husemann et al., 2012). And the intraspecific structure of riverine fish is usually clustered by river basins (Bartáková et al., 2015).

H. leucisclus often occur in high abundance in river and lake, because of high fertility, potentially reducing the amount of genetic drift through counteracting genetic differentiation

among populations (Tibbets and Dowling, 1996). Further, passive dispersal by flooding may reduce genetic differentiation, at least for riverine fish within river basins (Van Leeuwen et al., 2013). Though the distribution of freshwater fish is often limited by water area, long-distance dispersal may occur when fish or their larvae are transported passively, either by accidentally or on purpose by humans. Hence, the broadly distributed *H. leucisclus* tends to show high connectivity and gene flow mediated by strong adaptation and fecundity, as well as their aggregative characteristic.

However, despite multiple modes of passive dispersal, strong population divergence is often observed among local populations in the freshwater system (Husemann et al., 2012), either resulting of natural geographic isolation or anthropogenic fragmentation. In China, the rapid uplift of Tibet Plateau changed the trajectory of the river system in China, and ultimately leading to the formation of a flowing eastward drainage system (Clark et al., 2004). The Yangtze River was built between Oligocene and Miocene, but the unification of the upper and middle Yangtze River in the Three Gorges mountain region has happened between Late Pliocene and Early Pleistocene (Zheng et al., 2013). However, it is generally believed that the drainage pattern of the modern Yellow River was closely followed by the great uplift of the Tibet plateau once again, about 1.7 million years ago (Mya), but the ultimate formation of the Yellow River was at about 0.15 Mya, when Sanmenxia was completely cut through by Yellow River and went through to east (Li et al., 2001). The development of major river systems in the Pearl River Basin was derived from the Yanshan tectonic movement, the connectivity between the Yangtze River and the Pearl River system was believed before the early Pleistocene (Miao et al., 2004). However, the Pearl River basins and Yangtze River basins are divided by the Yunnan-Guizhou Plateau and Nanling Mountains, which hold back the cold weather from the north, forming the warmer climate and well-watered tropical land in the Pearl River basins in China. As a result, the distinct and different evolutionary histories of the river systems can be expected to affect current patterns of diversity of the widely distributed freshwater species.

In this study, we used the partial mitochondrial cytochrome b gene (cytb, mitochondrial gene) and the recombination activating gene 2 (RAG2, nuclear DNA) and combined the data of previous studies to estimate the phylogeographic pattern and evolutionary history of *H. leucisclus* across the three major river systems in China, and to evaluate the conservation status and strategy of this species.

MATERIALS AND METHODS

Samples and Sequence Data Preparation

Ten populations of *H. leucisclus* in the Yellow River basin and six populations in Huaihe River basin (**Supplementary File 1**) were collected between May 2017 and August 2018. The genomic DNA was extracted and the cytb was amplified following Fan and He (2014). We sequenced 345 individuals *de novo* (accession number in GenBank, ON315395-ON315739) and withdrew 725 sequences from the previous studies (Chen et al., 2017; Wang L. H. et al., 2021). A total of 1,070 cytb sequences from 59 localities across three major River systems in China were analyzed in this study (**Figure 1** and **Supplementary File 1**). In addition, the RAG2 gene was also amplified following Chen et al. (2017) and the new data (accession number in GenBank, ON315740-ON315840) included 101 individuals distributed in the Yellow River and Huaihe River basin. Another 45 RAG2 sequences were also retrieved from a previous study (Chen et al., 2017), and a total of 146 RAG2 sequences were analyzed (**Supplementary File 1**). Cycle-sequencing reactions were executed using BigDye terminator v. 3.1 and analyzed using an ABI-PRISM 3730 sequencer at Sangon Biotech (Shanghai, China). Both DNA strands were sequenced for both new cytb and RAG2 fragments in this study.

Phylogenetic Analysis

The new datasets generated and analyzed during this study are available in the GenBank (**Supplementary File 1**). All sequences were aligned by using MAFFT implemented in PhyloSuite v1.2.2 (Zhang et al., 2020). The aligned nucleotide sequences cytb and RAG2 were concatenated and used to construct the phylogenetic tree by using the maximum likelihood (ML) method implemented in RAxML v8.2.4 with 1,000 bootstrap replicates. Each gene was treated as unlinked. We were able to include all 1,070 individuals assayed since it is not required to have the same individuals analyzed for every genetic marker if partitions are considered unlinked. Bayesian inferences (BI) analysis was performed in MrBayes v3.2.6 using the Markov Chain Monte Carlo method with 10 million generations and sampling trees every 1,000 generations. The first 25% of trees were discarded as burn-in with the remaining trees being used for generating a consensus tree. The species of *Hemiculter bleekeri* and *Culter alburnus* (**Supplementary File 1**) were used as outgroups. The final trees were visualized in FIGTREE v1.4.4.

To further visualize haplotype diversity and distribution of *H. leucisclus*, we generated haplotype networks using 1,070 cytb sequences and colored each haplotype by the geographic region from which river system it was collected (**Supplementary File 1**). Another four haplotype networks in different river systems were also generated using cytb sequences. The haplotype networks were constructed using network v. 10[1] and applying the median-joining and maximum parsimony options. Four sublineages were found in this study, and the genetic diversity of each was calculated using DNASP v6.12.03.

Divergence Time Estimation

The divergence time was estimated using a molecular clock approach as implemented in BEAST v2.6.6. We used a reduced dataset of cytb: specimens for which one or six sequences from the same region were included. This approach resulted in 322 terminals—317 representatives of the 58 populations of *H. leucisclus* and five outgroup sequences (*H. bleekeri* and *C. alburnus*) (**Supplementary File 2**). As the lacking of fossil evidence, we first used a conservative approach by employing the strict-clock model from previous studies (Chen et al., 2017),

[1] www.fluxus-engineering.com/sharenet.htm

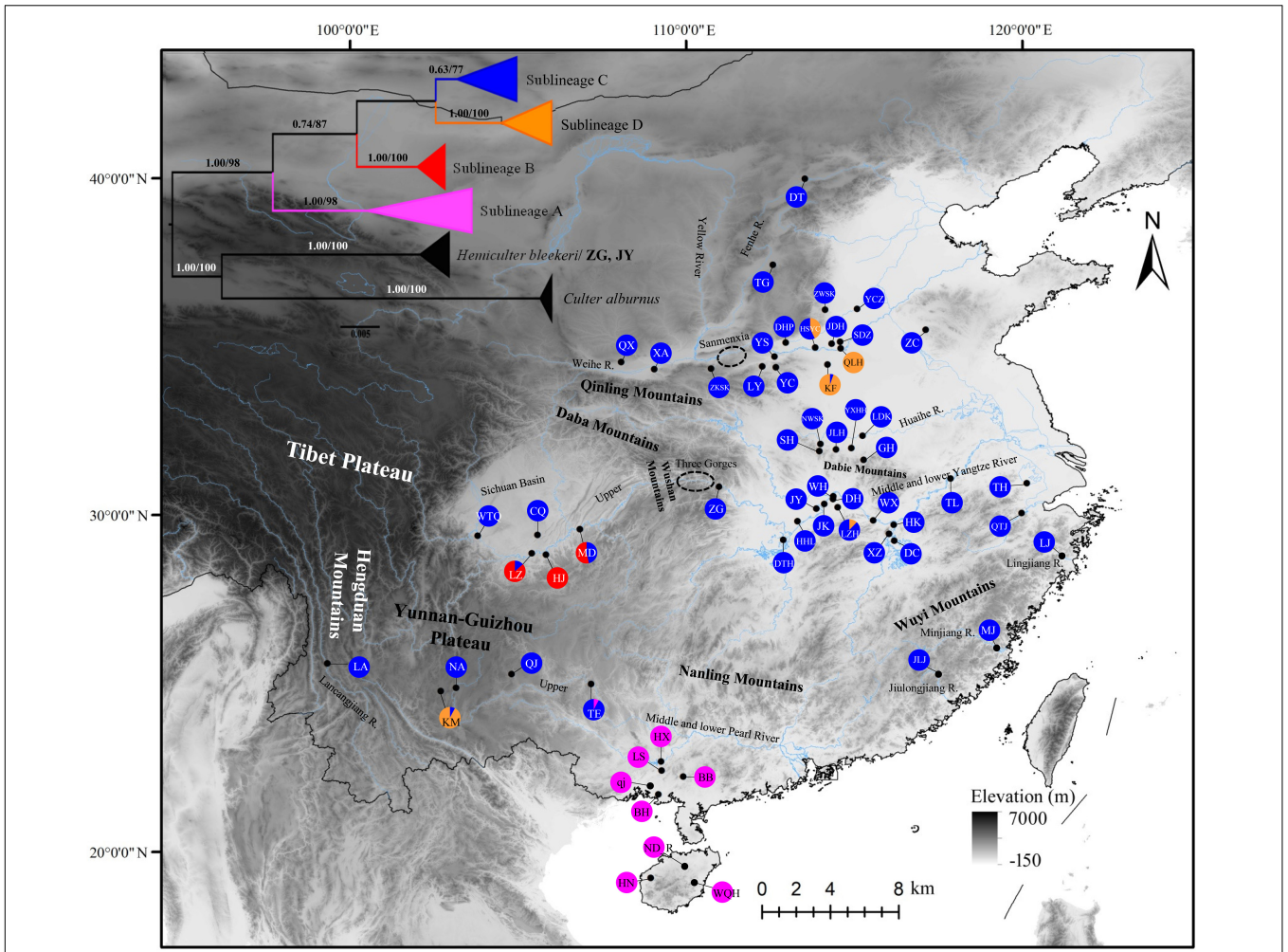


FIGURE 1 | Map of sampling locations of *H. leucisculus*, most locations in middle stream of Yellow River were cited from Wang L. H. et al., 2021, and all the locations in Yangtze River system and Pearl River system were cited from Chen et al. (2017) and Wang Y. Q. et al. (2021), and the phylogenetic tree based on Bayesian inference (values on branches indicate bootstrap proportions from a maximum likelihood analysis and Bayesian posterior probabilities). Localities are detailed in **Supplementary File 1**. Populations are presented as pie charts with slice sizes proportional to the frequency of the four lineages as identified by phylogenetic analysis. Colors correspond to the lineages (purple, sublineage A; red, sublineage B; blue, sublineage C; orange, and sublineage D).

a range of substitution rate of 1.00–2.00% per million years is often adopted and widely employed for cytb gene analysis in cyprinid fish (Durand et al., 2002; Ketmaier et al., 2004). The GTR + I + G model was implemented, and 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. Convergence was checked with TRACER v1.7.1. Posterior trees from the four runs were combined after removing the first 10% as burn-in in LogCombiner. The maximum credibility tree was created in TreeAnnotator available in the BEAST package. We performed another molecular clock analysis with similar settings but using a "relaxed clock model" with two calibration points. The first calibration point was the timing of separation between the upper Pearl River and Yangtze River during the early Pleistocene (2.5–1.7 Mya). The node was calibrated using a normal distribution with a mean at 2.1 Mya and a standard deviation of 0.2 Mya. According to the Cyprinidae fossils found in China, the recent cyprinid fauna of the East Asian mainland (Wang, 2005), including the Cultrins, was most likely established in the Pliocene, which was used as an additional calibration point for the common ancestor of *Culter* and *Hemiculter*.

Reconstruction of Ancestral Areas

We further performed a biogeographic reconstruction of ancestral areas for species of *H. leucisclus* in China using BioGeoBEARS in RASP v4.02. The analyses were conducted on a fully resolved topology from the BEAST analysis containing a subset of individuals which covered the range of four cytb mtDNA lineages of *H. leucisclus* and *H. bleekeri* (**Supplementary File 3**). Seven major geographical areas were defined based on the distribution of *H. leucisclus*: (A) the Yellow River basin, (B) the upper Yangtze River basin, (C) the middle and lower Yangtze River basin, (D) the Pearl

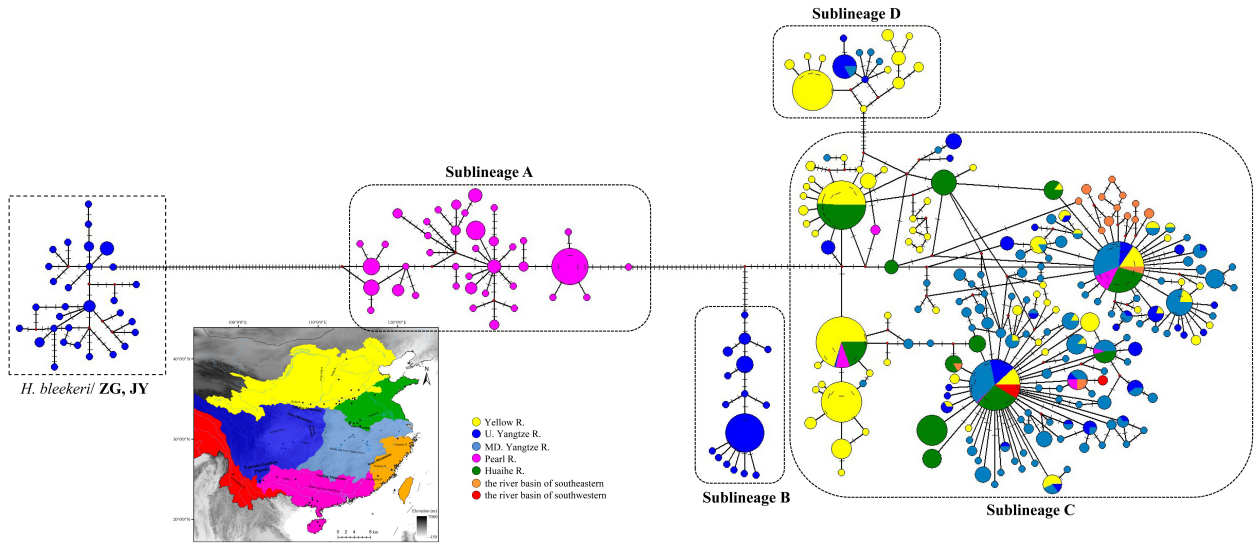


FIGURE 2 | Median-joining haplotype network based on 1,070 cytb sequences of *H. leucisculus* (see **Supplementary File 1**). The size of the circles represents haplotype frequency. Each connecting line represents a single nucleotide substitution, and each little short line represents a mutated position. Different river systems were represented in different colors, consistent with the colors of haplotype.

River basin, (E) the Huaihe River basin, (F) the river basin of southeastern, and (G) the river basin of southwestern. 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 by comparing Akaike's information criterion and likelihood−ratio tests, and DIVALIKE + *j* was chosen.

RESULTS

Phylogenetic Tree and Haplotype Network

The ML and BI tree showed a clear identification of four sublineages found in *H. leucisclus* (sublineage A, B, C, and D, **Figure 1**). However, most specimens of *H. leucisclus* collected from ZG and JY in the Yangtze River basin were classified as *H. bleekeri* both in the ML and BI tree. Furthermore, the median joining network of cytb showed that there were so many mutated positions between the haplotype collected from ZG and from other localities (**Figure 2**), indicating a very distant phylogenetic relationship. Our results suggested that the specimens from ZG and JY were likely misidentified as *H. leucisclus* in the previous study (Chen et al., 2017). In addition, four mtDNA lineages coinciding with the ML and BI trees were found in the haplotype network, and a significant phylogeographical pattern with two specific regions was found for *H. leucisclus*, one in the Pear River basin (sublineage A), and another in upper Yangtze River basin (sublineage B) (**Figure 2**). Sublineage C had a wide distribution across the six river systems and there were many shared haplotypes, while sublineage D had a relatively narrow distribution range in the Yangtze River and Yellow River basins. The genetic diversity indices showed a high level of genetic diversity in different sublineages, except for sublineage B (π = 0.00406, h = 0.87, **Supplementary Table 1**).

Divergence Time Estimation

We utilized a constant trait specific cytb Cyprinidae clock, specimens from different river systems form four distinctive clades. The divergence between *H. bleekeri* and *H. leucisclus* was at approximately 3.69 Mya. The divergence of sublineage A was dated at approximately 2.27 Mya, and the split time of sublineage B was about 1.42 Mya. Sublineage C diverged from sublineage at approximately 0.83 Mya (**Figure 3**). By applying a fossil calibrated molecular clock, we estimated the molecular diversification between *H. leucisclus* and its sister group of *H. bleekeri* was approximately 4.25 Mya. Sublieage A was estimated to have separated at about 2.57 Mya, followed by the divergence of sublineage B at about 1.62 Mya. Sublineage D separated from C at approximately 0.96 Mya (**Figure 3**).

Ancestral Range Reconstruction

The results of the ancestral range reconstruction are depicted in **Figure 4**, which showed that the most recent common ancestor (TMRCA) of *H. leucisclus* was mainly distributed in the Pearl River basin (node 35, D: 44.49%). There existed 20 dispersal and 11 vicariance events that occurred during the evolution of *H. leucisclus* (**Supplementary Tables 2**, **3**). TMRCA of *H. leucisclus* probably diverged into two distinct clades in the late Pliocene. One clade was just distributed in the middle and lower Pearl River basin, while the other clade further split into two separate sublineages with no dispersal and vicariance event.

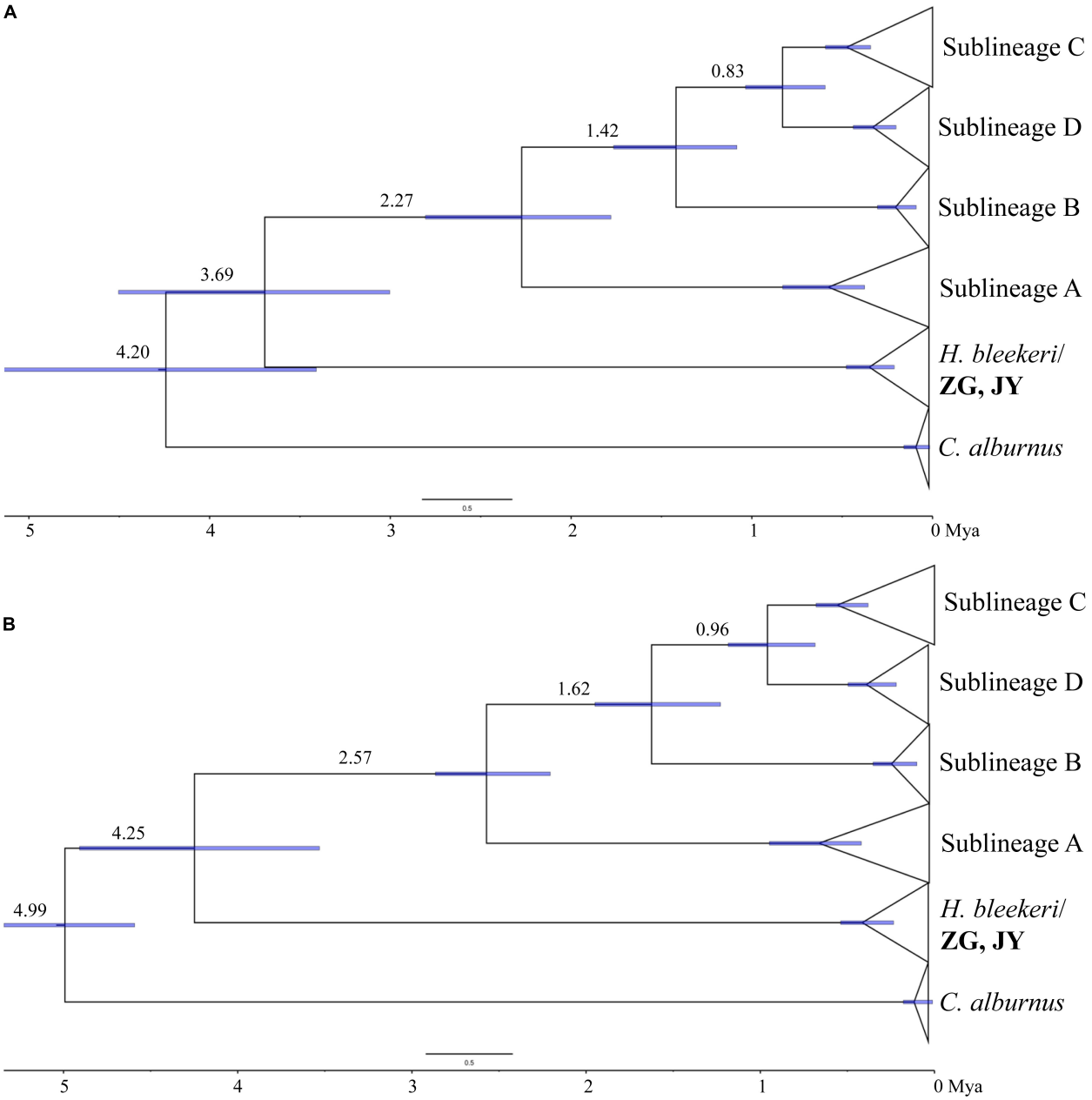


FIGURE 3 | Dated phylogeny **(A)** using a rate calibration; **(B)** using geological (red dot 1: 2.5–1.7 Mya) and fossil calibration points (red dot 2: 5.3–2.58 Mya) to estimate the divergence of *H. leucisculus*. Node ages are shown together with 95% highest posterior density bars indicating a range of age estimates.

DISCUSSION

The role of past climate fluctuations and geological events, playing in extant species diversification and distribution, is one of the main focuses of phylogenetics and biogeography (Rull, 2011). One consequence of drastic climatic change and geological events is the frequent reconfiguration of landscape. Meanwhile, spatial and geographical barriers prevent gene flow, which in turn affects the phylogeographic pattern (Tittes and Kane, 2014). However, rivers can act as dispersal corridors for fish, promoting or counteracting geographical isolation and diversification (Vernesi et al., 2016).

Our results identified four distinct sublineages for *H. leucisclus* across different river systems in China with a significant phylogeographic patterns. One of the sublineages, sublineage C, had a very wide distribution across all the river systems in this study, even distributed in many isolated drainage systems, including Lancangjiang River, the upper Pearl River, the Fenhe River, and Weihe River (in Yellow River system), Huaihe River, and the river basins of southeastern China (**Figure 1**), suggesting

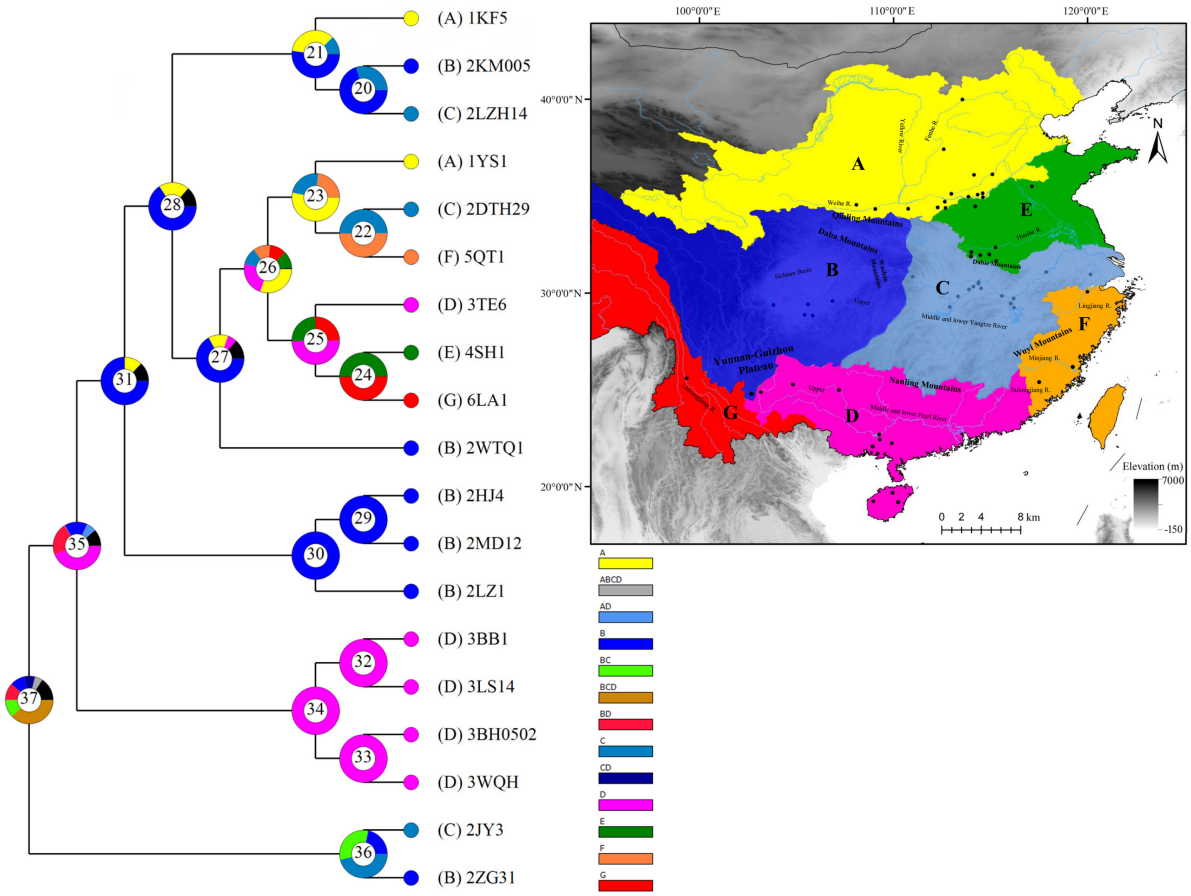


FIGURE 4 | Ancestral range reconstruction for the *H. leucisculus* across different river systems using 19 sequences of cytb. The colors of the charts correspond to the most likely ancestral areas inferred, and the black color means the unknown area. Letters represent the biogeographic regions for *H. leucisculus*: (A) the Yellow River basin, (B) the upper Yangtze River basin, (C) the middle and lower Yangtze River basin, (D) the Pearl River basin, (E) the Huaihe River basin, (F) the river basin of southeastern, and (G) the river basin of southwestern. The blue curves mean the river systems.

it's distributions did not completely match the river delimitation, as well as the geographic zone. The widely distributed sublineage C in different river systems could be attributed to their spatial proximity, which enables *H. leucisculus* populations from the Yangtze River basin to expand to nearby rivers easily by flooding in summer (Chen et al., 2017). However, the flooding cannot explain the strong admixture among far distant sampling locations across different river systems and even isolated drainage (**Figure 2**). Hence, anthropogenic translocations, especially the large-scale introduction of economically valuable fish half a century ago, offer alternatively possible explanations.

The formation of the Nanling Mountain Range could drive the separation of sublineage A from others (Wang L. H. et al., 2021), and the separation of sublineage B distributed in Sichuan Basin could be attributed to the isolation of populations by mountains barrier in the Southwestern and Central China (**Figure 1**). The Dabie Mountains-Eastern Qinling, Daba Mountains and Wushan Mountains have shaped the phylogeographic pattern of many other species, including *Sinopotamon acutum* (Fang et al., 2015), *Microtus fortis* (Gao et al., 2017), and *Hyalessa maculaticollis* (Liu et al., 2018). However, the populations from the Yellow River and Huaihe River basins had not yet formed independent lineage, although there are many watersheds existing in Qinling-Dabie Mountains. Furthermore, the populations from the middle and lower Pearl River basin and Hainan Island were unexpectedly found to have a close relationship. Considering the molecular data, the divergence time of *H. leucisclus* far postdated the formation of these mountains. For example, the ultimate formation of Nanling Mountain and the watershed between Yangtze River and Pearl River basin was dated to the Himalayan movement during the tertiary (Wang, 1993), suggesting that dispersal rather than vicariance has created the current distribution pattern of *H. leucisclus* in China.

Furthermore, the climate oscillations and geologic events during the quaternary could be a major driver of intraspecific divergence for *H. leucisclus*. Sublineage A with the highest nucleotide diversity (π = 0.00742, **Supplementary Table 1**) only existed in warm and moist tropical or subtropical areas, which provided important refugia for freshwater species during glaciations (Gao et al., 2012; Yan et al., 2013). Our results

suggested that *H. leucisclus* originated in the Pearl River basin at 4.25–3.69 Mya (**Figures 3A,B**). And our data are consistent with the conclusion of previous studies (Chen et al., 2017). And the divergence time of lineage A was at approximately 2.57–2.27 Mya, which fit well with the separation of Upper Yangtze River and Pearl River during Late Pliocene and Early Pleistocene (Miao et al., 2004). It was speculated that the active river capture and reverse events may have caused the connection of the Yangtze River and the upper Pearl River in some regions before the Early Pleistocene. Because the upper Yangtze River originally drained into the South China Sea through the Paleo-Red River until it was captured by the middle and lower reaches of the Yangtze River and reversed its route to flow into the East China Sea (Perdices et al., 2005). The resulting river capture ultimately generated the isolation between the Yangtze River and Pearl River systems, and in turn, likely had a dramatic effect on the evolution of many freshwater species (Gu et al., 2019). Furthermore, the separation of sublineage B at approximately 1.42–1.62 Mya exactly coincided with the timescale of the formation of the current Yangtze River (Li et al., 2001). It was reported that the unification of the upper and middle Yangtze River in the Three Gorges Mountain region occurred between the Late Pliocene and Early Pleistocene (Zheng et al., 2013). Hence, the connection of the upper and middle Yangtze River facilitated species dispersal from west to east, similar to what was observed in the frog *Quasipaa boulengeri* (Yan et al., 2013) and the freshwater snail *Bellamya* (Gu et al., 2019). Therefore, the establishment of the eastward-flowing the Yangtze River and the changes in river systems could have provided opportunities for dispersal and gene flow between closely related freshwater species, especially fish. It was surprising that sublineage D had a relatively narrow and discontinuous distribution in Yangtze River and Yellow River basins. The divergence time estimation between sublineage C and D was at 0.96–0.83 Mya, basically matched with the Dagu Glaciation (1.1∼0.9 Mya) in China (Jing and Liu, 1999). The result of repeated episodes of range contraction and expansion during glacial cycling has been hypothesized to be important driving forces of vicariant speciation or intraspecific divergence in many fish species (Chen et al., 2017; Li et al., 2020).

The evolutionary significant units have been proposed for the assessments of biodiversity and have a significant impact on conservation program (Niemiller et al., 2012), especially the recognition of phylogeographic patterns has important implications for the accurate assessment of the conservation of genetic resources. According to the distribution of the four sublineages in different river systems, two sublineages were found in the Yellow River basin, three sublineages in the Yangtze River basin, and two sublineages in the Pearl River basin (**Supplementary Figure 1**), suggesting that the genetic structure of *H. leucisclus* has not been completely changed by recent human-mediated translocations and flooding. Based on our new findings, the conservation of *H. leucisclus* should be reassessed, as the current conservation strategies are performed under the assumption of only one species and one gene pool. Especially for the sublineage B in a couple of rivers in the upper Yangtze River basin with very low nucleotide diversity. In a previous study, the morphological diversification found in *H. leucisculus* along climate gradients is primarily caused by the divergent selection pressures among different environments (Wang L. H. et al., 2021). Given these indications and molecular information, we advocate different protection strategies for *H. leucisculus* in different river systems. There should be two protection units for *H. leucisculus* in the Yellow River basin, three protection units for *H. leucisculus* in the Yangtze River basin, and two protection units for *H. leucisculus* in the Pearl River basin. While only one protection unit for *H. leucisculus* is necessary in Huaihe River basin. Furthermore, a thorough survey of wild populations of these lineages, including an assessment of their ecology, will be required for a more precise evaluation of their conservation status.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care Committee of Hunan Normal University.

AUTHOR CONTRIBUTIONS

QG conceived and designed the study. HZ and YS analyzed the data. HY, HZ, SL, and ZS carried out tissue dissection, DNA isolation, sample preparation for sequencing, and collect the mtDNA data from GenBank. QG and HZ drafted the manuscript. MW revised the language. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Natural Science Foundation of Hunan (Grant No. 2021JJ30442) and the National Natural Science Foundation of China (No. 31601851).

ACKNOWLEDGMENTS

We greatly appreciate the reviewers for their very helpful comments on previous versions of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.865089/full#supplementary-material

REFERENCES

Bartáková, V., Reichard, M., Blažek, R., Polačik, M., and Bryja, J. (2015). Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna. *J. Biogeogr.* 42, 1832–1844. doi: 10.1186/1471-2148-13-196

Chen, W. T., Schmidt, B. V., and He, S. P. (2017). The potential colonization histories of *Opsariichthys bidens* (Cyprinidae) in China using Bayesian binary MCMC analysis. *Gene* 676, 1–8. doi: 10.1016/j.gene.2018.07.019

Clark, M. K., Schoenbohm, L. M., Royden, L. H., Whipple, K. X., Burchfiel, B. C., Zhang, X., et al. (2004). Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics* 23, 1006–1029.

Durand, J. D., Tsigenopoulos, C. S., Unlu, E., and Berrebi, P. (2002). Phylogeny and biogeography of the family Cyprinidae in the Middle East inferred from cytochrome b DNA - evolutionary significance of this region. *Mol. Phylogenet. Evol.* 22, 91–100. doi: 10.1006/mpev.2001.1040

Fan, Q., and He, S. P. (2014). The pattern of upper and middle Yangtze drainages shapes the genetic structure and diversity of *Hemiculter leucisculus* revealed by mitochondrial DNA locus. *Acta Hydrobiol. Sin.* 38, 627–635.

Fang, F., Ji, Y., Zhao, Q., and Wang, Y. (2015). Phylogeography of the Chinese endemic freshwater crab *Sinopotamon acutum* (Brachyura, Potamidae). *Zool. Scr.* 44, 653–666. doi: 10.1111/zsc.12131

Gao, J., Yue, L. L., Jiang, X., Ni, L., Ashra, M. A., Zhou, Y., et al. (2017). Phylogeographic Patterns of *Microtus fortis* (Arvicolinae: Rodentia) in China Based on Mitochondrial DNA Sequences. *Pak. J. Zool.* 49, 1185–1195. doi: 10.17582/journal.pjz/2017.49.4.1185.1195

Gao, Y., Wang, S. Y., Luo, J., Murphy, R. W., Du, R., Wu, S., et al. (2012). Quaternary palaeoenvironmental oscillations drove the evolution of the Eurasian *Carassius auratus* Complex (Cypriniformes, Cyprinidae). *J. Biogeogr.* 39, 2264–2278. doi: 10.1111/j.1365-2699.2012.02755.x

Gu, Q. H., Husemann, M., Wu, H. H., Dong, J., Zhou, C. J., Wang, X. F., et al. (2019). Phylogeography of *Bellamya* (Mollusca: Gastropoda: Viviparidae) snails on different continents: contrasting patterns of diversification in China and East Africa. *BMC Evol. Biol.* 19:82. doi: 10.1186/s12862-019-1397-0

Husemann, M., Ray, J. W., King, R. S., Hooser, E. A., and Danley, P. D. (2012). Comparative biogeography reveals differences in population genetic structure of five species of stream fishes. *Biol. J. Linn. Soc.* 107, 867–885. doi: 10.1111/j.1095-8312.2012.01973.x

Jing, C. R., and Liu, H. P. (1999). On the glacial and Interglacial stages in Quaternary of China. *J. Chengdu Univ. Technol.* 26, 97–100.

Ketmaier, V., Bianco, P. G., Cobollia, M., Krivokapić, M., Caniglia, R., and De Matthaeis, E. (2004). Molecular phylogeny of two lineages of Leuciscinae cyprinids (*Telestes* and *Scardinius*) from the peri-Mediterranean area based on cytochrome b data. *Mol. Phylogenet. Evol.* 32, 1061–1071. doi: 10.1016/j.ympev.2004.04.008

Leclerc, E., Mailhot, Y., Mingelbier, M., and Bernatchez, L. (2008). The landscape genetics of yellow perch (*Perca flavescens*) in a large fluvial ecosystem. *Mol. Ecol.* 17, 1702–1717. doi: 10.1111/j.1365-294X.2008.03710.x

Li, C., Jiang, S. Y., Schneider, K., Jin, J., Lin, H., Wang, J., et al. (2020). Cryptic species in White Cloud Mountain minnow, Tanichthys albonubes: Taxonomic and conservation implications. *Mol. Phylogenet. Ecol.* 153:106950. doi: 10.1016/j.ympev.2020.106950

Li, J. J., Fang, X. M., Pan, B. T., Zhijun, Z., and Yougui, S. (2001). Late Cenozoic intensive uplift of Qinghai-Xizang Plateau and its impacts on environments in surrounding area. *Quat. Sci.* 21, 381–391.

Liu, Y., Qiu, Y., Wang, X., Yang, H., Hayashi, M., Wei, C., et al. (2018). Morphological variation, genetic differentiation and phylogeography of the East Asia cicada *Hyalessa maculaticollis* (Hemiptera: Cicadidae). *Syst. Entomol.* 43, 308–329. doi: 10.1111/syen.12276

Manel, S., Schwartz, M. K., Luikart, G., and Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18, 189–197. doi: 10.1016/s0169-5347(03)00008-9

Miao, W., Yu, Y. H., Shen, Y. F., and Zhang, X. (2004). Intraspecific phylogeography of *Carchesium polypinum* (Peritrichia, Ciliophora) from China, inferred from 18S-ITS1-5.8S ribosomal DNA. *Sci. China Ser. C Life Sci.* 47, 11–17. doi: 10.1360/02yc0157

Niemiller, M. L., Near, T. J., and Fitzpatrick, B. M. (2012). Delimiting species using multilocus data: diagnosing cryptic diversity in the southern cavefish, *Typhlichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution* 66, 846–866. doi: 10.1111/j.1558-5646.2011.01480.x

Perdices, A., Sayanda, D., and Coelho, M. M. (2005). Mitochondrial diversity of *Opsariichthys bidens* (Teleostei, Cyprinidae) in three Chinese drainages. *Mol. Phylogenet. Evol.* 37, 920–927. doi: 10.1016/j.ympev.2005.04.020

Rull, V. (2011). Neotropical biodiversity: timing and potential drivers. *Trends Ecol. Evol.* 26, 508–513. doi: 10.1016/j.tree.2011.05.011

Schultheiß, R., Van Bocxlaer, B., Riedel, F., von Rintelen, T., and Albrecht, C. (2014). Disjunct distributions of freshwater snails testify to a central role of the Congo system in shaping biogeographical patterns in Africa. *BMC Evol. Biol.* 14:42. doi: 10.1186/1471-2148-14-42

Tibbets, C. A., and Dowling, T. E. (1996). Effects of intrinsic and extrinsic factors on population fragmentation in three species of North American minnows (Teleostei: Cyprinidae). *Evolution* 50, 1280–1292. doi: 10.1111/j.1558-5646.1996.tb02368.x

Tittes, S., and Kane, N. C. (2014). The genomics of adaptation, divergence and speciation: a congealing theory. *Mol. Ecol.* 23, 3938–3940. doi: 10.1111/mec.12855

Van Leeuwen, C. H. A., Huig, N., Van Alen, T. A., and Van der Velde, G. (2013). How did this snail get here? Several dispersal vectors inferred for an aquatic invasive species. *Freshw. Biol.* 58, 88–99. doi: 10.1111/fwb.12041

Vernesi, C., Hoban, S. M., Pecchioli, E., Crestanello, B., Bertorelle, G., Rosà, R., et al. (2016). Ecology, environment and evolutionary history influence genetic structure in five mammal species from the Italian Alps. *Biol. J. Linn. Soc.* 117, 428–446. doi: 10.1111/bij.12651

Wang, C. L. (1993). Formation and evolution of Nanling Range. *Trop. Geomorphol.* 14, 46–52.

Wang, L. H., Zhu, L., Tang, K., Liu, M., Xue, X., Wang, G., et al. (2021). Population genetic structur of sharpbelly *Hemiculter leucisculus* (Basilesky, 1855) and morphological diversification along climate gradients in China. *Ecol. Evol.* 11, 6798–6813. doi: 10.1002/ece3.7528

Wang, X. Z. (2005). *Molecular Phylogeny of the East Asian Cyprinids (Pisces: Cypriniformes)*. Wuhan: Institute of Hydrobiology, Chinese Academy of Sciences.

Wang, Y. Q., Feijó, A., Cheng, J. L., Xia, L., Wen, Z., Ge, D., et al. (2021). Ring distribution patterns—diversification or speciation? Comparative phylogeography of two small mammals in the mountains surrounding the Sichuan Basin. *Mol. Ecol.* 30, 2641–2658. doi: 10.1111/mec.15913

Yan, F., Zhou, W. W., Zhao, H. T., Yuan, Z., Wang, Y., Jiang, K., et al. (2013). Geological events play a larger role than Pleistocene climatic fluctuations in driving the genetic structure of *Quasipaa boulengeri* (Anura: Dicroglossidae). *Mol Ecol.* 22, 1120–1133. doi: 10.1111/mec.12153

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096

Zheng, H. B., Wang, P., He, M. Y., Luo, C., Huang, X. T., Jia, J. T., et al. (2013). Timing of the Establishment of the East-Flowing Yangtze River and Tectonic-Geomorphic Implications. *Quat. Sci.* 33, 621–630.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

BRIEF RESEARCH REPORT
published: 13 June 2022
doi: 10.3389/fgene.2022.908367

Analysis of Population Genetic Diversity and Genetic Structure of *Schizothorax biddulphi* Based on 20 Newly Developed SSR Markers

Zhulan Nie[1,2,3,4], *Yongli Ren*[1,2,3], *Lirong Zhang*[2,3,4], *Rui Ge*[2,3,4] *and Jie Wei*[1,2,3,4]*

[1]College of Animal Science and Technology, Tarim University, Alar, China, [2]Key Laboratory of Tarim Animal Husbandry and Science Technology of Xinjiang Production and Construction Corps., Alar, China, [3]State Key Laboratory Breeding Base of Protection and Utilization of Biological Resources in Tarim Basin, Xinjiang Production and Construction Crops and Ministry of Science and Technology, Alar, China, [4]College of Life Science, Tarim University, Alar, China

To protect the germplasm resources of *Schizothorax biddulphi*, we developed and used 20 pairs of polymorphic microsatellite primers to analyze the genetic diversity and structure of populations. A total of 126 samples were collected from the Qarqan River (CEC), Kizil River (KZL), and Aksu River (AKS) in Xinjiang, China. The results showed that 380 alleles were detected in 20 pairs of primers and the average number of alleles was 19.0. The effective allele numbers and Nei's gene diversity ranged from 1.1499 to 1.1630 and 0.0962 to 0.1136, respectively. The Shannon index range suggested low levels of genetic diversity in all populations. The genetic distance between the CEC and AKS populations was the largest, and the genetic similarity was the smallest. There was a significant genetic differentiation between CEC and the other two populations. The UPGMA clustering tree was constructed based on population genetic distance, and the clustering tree constructed by individuals showed that the AKS population and KZL population were clustered together, and the CEC population was clustered separately. Also, the group structure analysis also got the same result. It can be seen that although the three populations of *S. biddulphi* do not have high genetic diversity, the differentiation between the populations was high and the gene flow was limited, especially the differentiation between the CEC population and the other two populations. This study not only provided genetic markers for the research of *S. biddulphi* but the results of this study also suggested the need for enhanced management of *S. biddulphi* populations.

Keywords: ***Schizothorax biddulphi*, microsatellite, genetic diversity, genetic structure, genome survey**

OPEN ACCESS

Edited by:
Cong Zeng,
Shanghai Jiao Tong University, China

Reviewed by:
Qianhong Gu,
Hunan Normal University, China
Zongjun Du,
Sichuan Agricultural University, China

***Correspondence:**
Jie Wei
weijiedky@126.com

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *30 March 2022*
Accepted: *02 May 2022*
Published: *13 June 2022*

Citation:
Nie Z, Ren Y, Zhang L, Ge R and Wei J (2022) Analysis of Population Genetic Diversity and Genetic Structure of Schizothorax biddulphi Based on 20 Newly Developed SSR Markers. Front. Genet. 13:908367. doi: 10.3389/fgene.2022.908367

INTRODUCTION

Schizothorax biddulphi, in the order Cypriniformes and family Cyprinidae, is a symbolic species distributed only in the Tarim River system of Xinjiang Uygur Autonomous Region, China (Le and Chen, 1998; Luo et al., 2012). Due to its slow growth, late sexual maturity, low fecundity, and higher requirements for living environment, coupled with the destruction of the ecological environment and human interference in recent years, the habitat of *S. biddulphi* has been destroyed and continuously reduced, and the population has been drastically reduced (Nie et al., 2014). Therefore, *S. biddulphi* was listed in the China Red Data Book of Endangered Animals–Fish Volume in 1998 (Le and Chen,

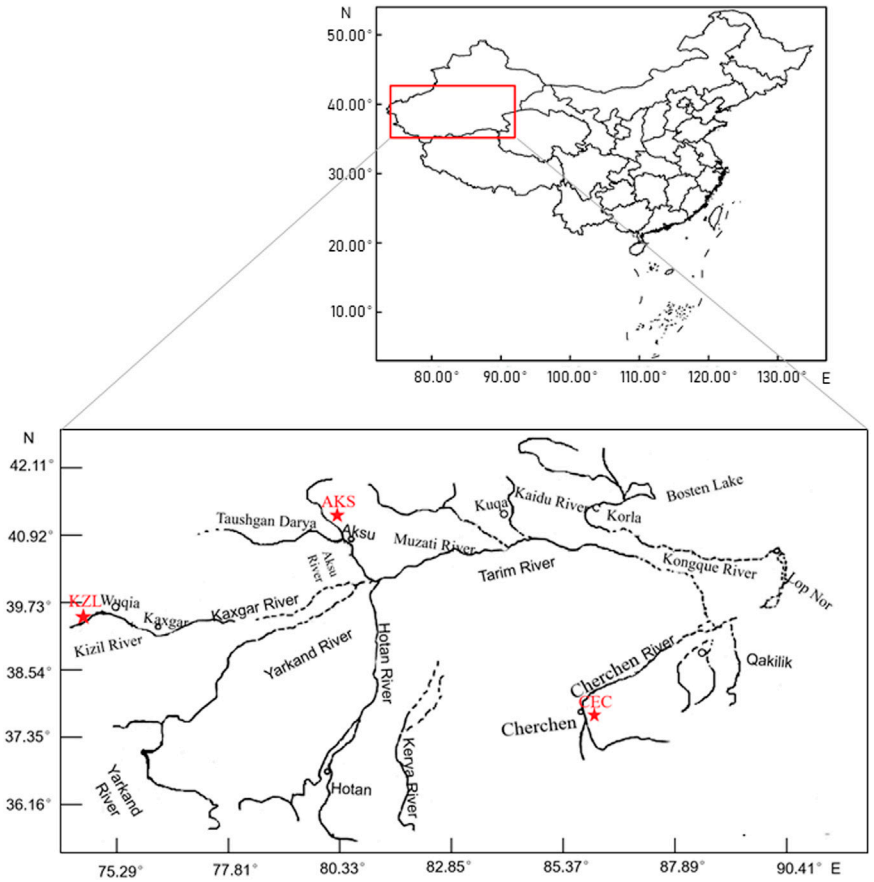


FIGURE 1 | Population sampling sites of *Schizothorax biddulphi*. ★: each sampling point.

1998) and in the Xinjiang Provincial Second-Class Protected Animals in 2004. In 2020, *S. biddulphi* upgraded to the national secondary protection fish of China.

Microsatellite DNA, also known as simple sequence repeats (SSRs) and short tandem repeats (STRs), is a nucleotide sequence consisting of 1–6 nucleotides in the repeat motif (Tautz and Renz, 1984). SSR has the advantages of large number, wide distribution, high polymorphism, high conservation, easy detection, high repeatability, and co-dominant characteristics, so it is widely distributed in eukaryotic genomes as a molecular marker (Li et al., 2008; Mohanty et al., 2016). SSR plays an important role in biological genetics research and has been increasingly used in conservation genetics research of endangered animals that lack genetic information, such as Rhesus macaque (*Macaca mulatta*), Asian elephant (*Elephas maximus*), Arabian oryx (*Oryx leucoryx*), and land snail (*Satsuma myomphala*), to guide the management and conservation of endangered animals (Li et al., 2012; Kang et al., 2017; Elmeer et al., 2019; Marasinghe et al., 2021; Qiu et al., 2021).

Most genetic diversity assessment methods are only applicable to diploid organisms, but studies showed that schizothoracine was polyploidy, mainly tetraploid, hexaploid, and octaploid (Wang et al., 2016; Guo et al., 2019). However, in most of the existing studies on *S. biddulphi*, the complexity of polyploidy was ignored, and the analyses followed the diploid approaches (Gong et al., 2012; Luo et al., 2012). This neglect might not explain population genetic diversity and genetic differentiation among populations more accurately. To overcome it, the most common solution was to convert microsatellite's polyploid genotypes to pseudodiploid genotypes, which allowed the use of many popular methods (Wang and Scribner, 2014; Liu et al., 2017). A few studies attempted to develop microsatellite primers for *S. biddulphi* to analyze genetic differences in populations based on polyploidy information, but the available markers were still very limited (Yang et al., 2014), and the polymorphism of these markers was low, which is difficult to revealing genetic differentiation among populations. Therefore, in this study, we wanted to develop more high polymorphic microsatellites for the research and population management of *S. biddulphi*. Also, three populations from the Tarim River Basin were investigated to provide support for the germplasm resource conservation of *S. biddulphi*.

MATERIALS AND METHODS

Materials

From July 2017 to July 2019, 126 tail fin samples were collected from 3 populations, including 44 from Cherchen River (CEC; 37°42′37″N, 85°41′45″E), 69 from Kizil River (KZL; 39°31′12″N,

TABLE 1 | 20 microsatellite locus characteristics of *Schizothorax biddulphi*.

Locus	Primer sequence (5′-3′)	Repeat motif	Tm (°C)	Product size (bp)	Accession
T43	F:TGTATGAAAGCACAATGGG R:GTGTAGTTTTTAGCGGCA	$(TG)_5(TA)_6$	48.0	164–191	MT211579
T90	F:TTATGATTTTCTTTCGTC R:TCAACATTCTCTCTTTTT	$(TA)_{12}$	49.0	339–349	MT211580
T136	F:GCTCTTGCTTTCTCTGGG R:GATTTCTCGTGCTGTCATT	$(AC)_{11}$	50.8	316–358	MT211581
T144	F:TGAGGTTTGGTGTCTGTG R:ATCCATCTGTCCGTCTGT	$(ATGG)_7$	56.6	91–154	MT211582
T166	F:GGTGTCATCTGTTTTGTGG R:CTGTATCCCTGGTTAGCG	$(GT)_5$	50.8	217–262	MT211583
T175	F:GTATTCCTGTATCTCACCTG R:TCTCATCCTCTTTCGTTC	$(TG)_6$	48.0	239–322	MT211584
T218	F:TGCTGGCAGAGAATGAATGT R:TTGGTTGATGTCAGAGGTTG	$(ACT)_{11}$	53.5	341–405	MT211585
T227	F:CTTTTTGTTGCTGAGGTGTT R:TACTGTGAGGGATTTGTCGT	$(CTA)_{16}$	49.0	272–372	MT211586
T229	F:GTAAAACCAACAGCACAGCC R:GGGACAAGGGGAAAAAACTA	$(CA)_{22}$	48.0	220–314	MT211587
T230	F:TTTCATCCTTCTGCCACCAC R:ATCCGCTAATCATAAACACC	$(AC)_{16}$	48.0	276–318	MT211588
T231	F:TATTACAAAACGAATGGAGG R:AAGAAACAAAGGTAGATCAA	$(AC)_{25}$	56.6	311–392	MT211589
T239	F:AAAGAAGGAAGAAAAGCAGC R:GGGGAAGGAAAAAAAGTGAC	$(TC)_{20}$	49.0	197–254	MT211590
T246	F:GTATCTGGTCCATCTGCTCC R:TCTGTCATGTTCTTTTGCTTT	$(GA)_{15}$	56.6	220–240	MT211591
T255	F:CCTGAACAAATAAAACACCA R:ACTCAAGGATAGACAAAACAT	$(TG)_{20}$	48.0	257–285	MT211592
T259	F:TGGGTTTTAGTATTTTTGTC R:ACTTTTTCTGTCTCGTTTCA	$(TA)_{20}$	48.0	351–372	MT211593
T260	F:CTTTCAAACAGACAAGAGGA R:AAACAGTTCAAGATGTAAATAAT	$(GA)_{22}$	48.0	218–258	MT211594
T269	F:CAAACAAACCTCTACCCTCC R:AAACAGTTCAAGATGTAAATAAT	$(CT)_{19}$	53.5	160–341	MT211595
T272	F:TATGTTATTGATGTGTTGATT R:GTGCACCCTTTAGTTGTTAG	$(AATG)_{14}$	56.6	253–301	MT211596
T277	F:AATGCTTTCCTTTCTCAGCT R:GTCCCTATTTTTTGTGGTTC	$(ATCT)_{15}$	48.0	281–380	MT211597
T278	F:AGCATCTTCTCATACATTTT R:GTGTCTTTATTTCTTGTCATT	$(ATCT)_{23}$	48.0	188–285	MT211598

75°14′20″E), and 13 from Akesu River (AKS; 40°14′4″N, 80°5′36″E), Xinjiang, China (**Figure 1**). All samples were stored in absolute alcohol and brought back to the laboratory and stored at −20°C. Total genomic DNA was extracted from the tail fin using the TIANamp Genomic DNA Kit (Tiangen, Beijing, China) following the manufacturer's instructions. The concentration and quality of DNA were estimated using a DeNovix DS-11 spectrophotometer (DeNovix, Wilmington, United States).

Genome Survey, SSR Locus Development, and Primer Design

The samples for the genome survey were randomly selected from 3 *S. biddulphi* from Cherchen River. Five DNA libraries of 270 bp(2), 350 bp(2), and 500 bp(1) insert size were constructed and sequenced by Illumina Hi-Seq 2000.

The Perl script MIcroSAtellite (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html) was used for genic-SSR markers with perfect repeat units of 2–6 nucleotides. The minimum SSR length criteria were defined as six reiterations for dinucleotide and five reiterations for other repeat units. About 288 microsatellite loci were selected, of which 288 with enough flanking sequence were used for primer design using Primer Premier 5.0 (Premier Biosoft International).

All SSR primer pairs were initially screened for a sample set comprising 42 randomly selected specimens, optimized for reproducible amplification using standard PCR conditions with annealing temperatures altered according to the primer sequences. By performing electrophoresis on an 8% non-denaturation polyacrylamide gel, 20 of 30 loci were successfully amplified and shown to be polymorphic, which were deposited in the GenBank (Accession Nos. MT211579–MT211598).

SSR Genotyping

The forward primers for the 20 SSR loci were labeled with fluorescent dyes (6-FAM, HEX, TAMRA, and ROX). PCR amplification was carried out using genomic DNA extracted from 126 *S. biddulphi*. Each 50 μL PCR contained 2 μL genomic DNA, 25 μL RTaq DNA Premix, 1 μL forward and reverse primers, and 21 μL double-distilled water. The thermocycling conditions were as follows: initial denaturation for 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at the annealing temperature (**Table 1**), and 40 s at 72°C, and finally 10 min at 72°C. The PCR products were separated by capillary electrophoresis using an ABI3730xl DNA analyzer (Applied Biosystems) after confirmation of PCR amplification on a 1.5% agarose gel. The genotyping of polymorphism microsatellite locus analysis was performed using the software GeneMapper v4.0 (Applied Biosystems).

Analysis of Population Genetic Diversity and Genetic Structure

This study referred to the method of Zhu et al. (2002) to analyze the electrophoretic patterns. The polymorphic bands amplified by each primer were counted, a band at the same migration position was marked as "1," and no band was marked as "0." According to the capillary electrophoresis pattern analyzed by Gene mapper 4.1, the standard for adopting the data is that the signal value is above 400, there is no interference of other spurious peaks, and the peak shapes from the same site are all similar. Finally, the original data matrix is established. Data editing and format conversion were performed in GenAlEx v 6.501 software.

Population genetic parameters such as the polymorphic percentage, allele number (N_a), effective number of alleles (N_e), Nei's genetic distance (H), and Shannon's Information index (I) were analyzed using GenePop v 4.14. Poptree V2 was used to establish an NJ clustering tree representing population genetic relationships based on allele frequency (Takezaki et al., 2010). Pcoa-genalex V6.5 analyzed genetic relationships among populations based on genetic distance (Smouse and Peakall, 2012). Structure 2.3.4 (http://pritch. bsd. uchicago. edu/) was used to analyze the genetic structure of the population, and the mixed model and frequency correlation model were selected. Online software Structure Harvest was used to estimate the theoretical population number according to the ΔK algorithm (Evanno et al., 2005).

RESULTS

Genome Survey in *Schizothorax biddulphi*

The genomic DNA of *S. biddulphi* was used to construct two 270 bp libraries, two 350 bp libraries, and one 500 bp library, which were sequenced and filtered to obtain 165.34 Gb of high-quality data. The total sequencing depth was about 174 ×, the Q20 ratio of sequencing data was above 95.63%, and the Q30 ratio was above 90.48% (**Supplementary Table SA1**).

The corresponding Kmer depth of the main peak is 120. There are three peaks in Kmer, the main peak, the 1/2 peak, and the 1/4 peak, suggesting that the species may be tetraploid (**Figure 2**). The total number of Kmer obtained from the sequencing data was 11,754,486,585. After removing the Kmer with abnormal depth, a total of 114,172,843,993 Kmer were used for genome length estimation, and the calculated genome length was about 947.30 Mbp. According to the Kmer distribution, the content of repeated sequences was estimated to be about 39.45%, and the heterozygosity was estimated to be about 0.81%. Therefore, the genome of this species belongs to a complex genome with high heterozygosity. Scaffold N50 was about 3.84 Kb and Contig N50 was about 869 bp (**Supplementary Table SA2**).

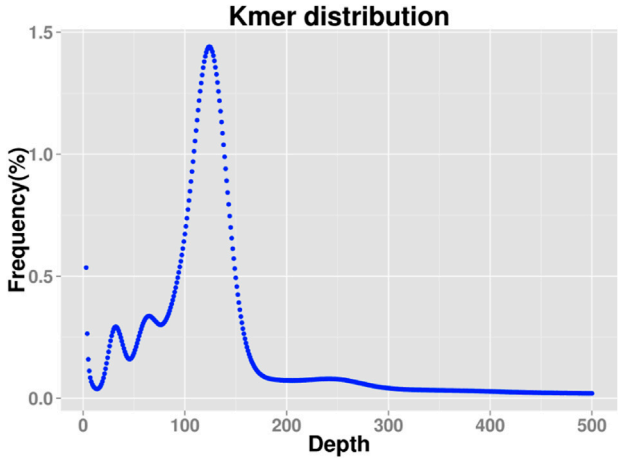


FIGURE 2 | 19 K-mer analysis for estimating the genome size of *Schizothorax biddulphi*.

Frequency Distribution of Different Types of SSR Markers

A total of 558,993 unigenes with a total nucleotide number of 1,125,446,683 were obtained from the genomic data of *S. biddulphi*. A total of 743,118 SSR sequences were detected, and 155,018 sequences contained more than 1 SSR (**Supplementary Table SA3**). The microsatellite-type richness of the detected *S. biddulphi* genome was high, and di-nucleotide repeat unit content was the most (42.74%), followed by mono-nucleotide repeat units (39.07%), and the content of tri-nucleotide (9.09%) and tetra-nucleotide (7.94%) repeat units was similar, the content of penta-nucleotide (0.92%) and hexa-nucleotide (0.24%) repeat units was less. The distribution density of the six nucleotide repeat types was positively correlated with the corresponding SSR content (**Table 2**).

S. biddulphi microsatellites include 4 different mono-nucleotide repeat microsatellite types, 12 different di-nucleotide repeat microsatellite types, 59 different tri-nucleotide repeat microsatellite types, 215 different tetra-nucleotide repeat microsatellite types, 582 different penta-nucleotide repeat microsatellite types, and 418 different hexa-nucleotide repeat microsatellite types. Among the mono-nucleotide repeats, A/T was the main base composition (93.23%). AC/TG (30.02%) was more in di-nucleotide repeat

TABLE 2 | Distribution characteristics of SSR motifs in the genome of *Schizothorax biddulphi*.

Repeat type	SSR count	Percent	Base type	Main repeat motif	Density of SSR (per/Mb)
Mono-nucleotide	290,331	39.07	4	A/T,C/G	204.99
Di-nucleotide	317,627	42.74	12	AC/TG,CA/GT,TA/AT,TC/AG,GA/CT	133.01
Tri-nucleotide	67,536	9.09	59	AAT/TTA,ATT/TA, ATA/TAT,AAC/TTG	39.37
Tetra-nucleotide	59,001	7.94	7.94	TCTA/AGAT, GATA/CTAT, AGAC/TCTG	18.36
Penta-nucleotide	6,862	0.92	582	TATTA/ATAAT, AAAAT/TTTTA	3.64
Hexa-nucleotide	1,761	0.24	418	GTGTGA/CACACT, ATATAC/TATATG	0.51
Total	743,118	100	1290	—	399.88

TABLE 3 | 20 microsatellite locus genetic variation of *S. biddulphi*.

Locus	N_a	N_e	*H*	*I*
T43	7	1.5802	0.3425	0.5137
T90	7	1.3942	0.2281	0.3508
T136	11	1.2636	0.1788	0.2994
T144	17	1.1913	0.1324	0.2322
T166	5	1.5804	0.3449	0.5188
T175	8	1.5004	0.2901	0.4441
T218	11	1.1471	0.117	0.2181
T227	25	1.1217	0.0925	0.1754
T229	21	1.1549	0.1155	0.2111
T230	8	1.2647	0.1646	0.2685
T231	39	1.0946	0.0789	0.1593
T239	26	1.1727	0.1246	0.2243
T246	11	1.3715	0.2221	0.3488
T255	12	1.1983	0.1428	0.2522
T259	17	1.1857	0.1356	0.2437
T260	24	1.2362	0.1641	0.2801
T269	56	1.0595	0.054	0.1206
T272	17	1.2659	0.1784	0.2966
T277	34	1.1066	0.0919	0.1838
T278	24	1.1449	0.1081	0.2046

TABLE 4 | Genetic diversity parameters of the six populations of *Schizothorax biddulphi*.

Population	Polymorphic Percentage	N_e	*H*	*I*
CEC	61.32	1.1615	0.1136	0.1936
KZL	66.84	1.1630	0.1084	0.1841
AKS	35.53	1.1499	0.0962	0.1529

TABLE 5 | Gene flow N_m (above diagonal) and F_{st} values for pairwise comparison (blow diagonal) among the three populations of *Schizothorax biddulphi*.

	CEC	KZL	AKS
CEC	—	0.897	0.866
KZL	0.218**	—	2.086
AKS	0.224**	0.107	—

** $p< 0.001$.

SSRs. AAT/TTA (29.44%) and ATT/TTA (23.57%) had high frequency of tri-nucleotide repeats. TCTA/AGAT and GATA/CTAT appeared more frequently in tetra-nucleotide repeats. TATTA/ATAAT and AAAAT/TTTTA appeared more frequently in penta-nucleotide repeats. GTGTGA/CACACT and ATATAC/TATATG appeared more frequently in hexa-nucleotide repeat microsatellite.

Polymorphism of SSR Markers

Among the 288 pairs of primers, 30 pairs of primers with high polymorphism were screened by agarose gel electrophoresis and 8% non-denaturing polyacrylamide gel electrophoresis. According to the detection results of capillary electrophoresis, 20 pairs of SSR primers with high polymorphism were screened out from the above 30 pairs of primers (**Table 1**). The polymorphic parameters of the 20 primer pairs are shown in **Supplementary Table SA4**.

Assessment of Genetic Diversity

A total of 380 alleles (N_a) were detected, and the number of alleles per SSR locus ranged from 5 (T166) to 56 (T269), with an average of 19 alleles. The number of effective alleles (N_e) ranged from 1.0595 (T269) to 1.5804 (T166), and the average number was 1.2517. Nei's genetic distance (*H*) ranged from 0.0540 (T269) to 0.3449 (T166), with an average of 0.1635. Shannon's Information index (*I*) for each SSR locus ranged from 0.1206 (T269) to 0.5188 (T166), with an average of 0.2773 (**Table 3**). In terms of population genetic diversity parameters, *H* and *I* were consistent, showing that the CEC was the highest and the AKS was the lowest (**Table 4**).

The range of genetic differentiation index and gene flow was 0.107–0.224 and 0.866–2.086, respectively. The genetic differentiation index among CEC, AKS, and KZL was great genetic differentiation ($0.15 < F_{st} < 0.25$), and there was moderate genetic differentiation between AKS and KZL ($0.05 < F_{st} < 0.15$). The gene flow between AKS and KZL was the highest, while the gene flow between CEC and the other two populations was low (**Table 5**).

The range of Nei's genetic distance and genetic identity between every two populations was 0.021–0.044 and 0.957–0.980, respectively. The genetic distance of Nei's between CEC and AKS was the longest (0.044), and the genetic identity was the least. The Nei's genetic distance between AKS and KZL was the closest (0.021), and the genetic identity was the largest (**Table 6**).

TABLE 6 | Nei's genetic distance D (above diagonal) and genetic identity (blow diagonal) among the three populations of *Schizothorax biddulphi*.

	CEC	KZL	AKS
CEC	—	0.038	0.044
KZL	0.962	—	0.021
AKS	0.957	0.980	—

Genetic Structure and Genetic Relationship Among *Schizothorax biddulphi* Populations

Cluster analysis was performed on an individual (**Figure 3**) and population (**Figure 4**) of *S. biddulphi*. The results showed that 126 samples of *S. biddulphi* were roughly divided into two clusters, with AKS and KZL individuals clustered into one cluster and CEC individuals clustered into another cluster. At the same time, the UPGMA tree constructed according to population genetic distance was consistent with the clustering tree constructed by individuals, the AKS and KZL clustered into a cluster, and the CEC clustered into a single cluster.

AMOVA results showed that 20% of the genetic variation of *S. biddulphi* came from inter-population and 80% of the genetic variation came from intra-population (**Table 7**). Therefore, the genetic variation within populations was larger than that between populations.

Division of the dataset into two subclusters ($K = 2$) produced the best assignment of individuals to subclusters, and subsequent analysis of each subset determined the appropriate number of populations within each subset (**Figure 5A**). There are two gene banks in the three

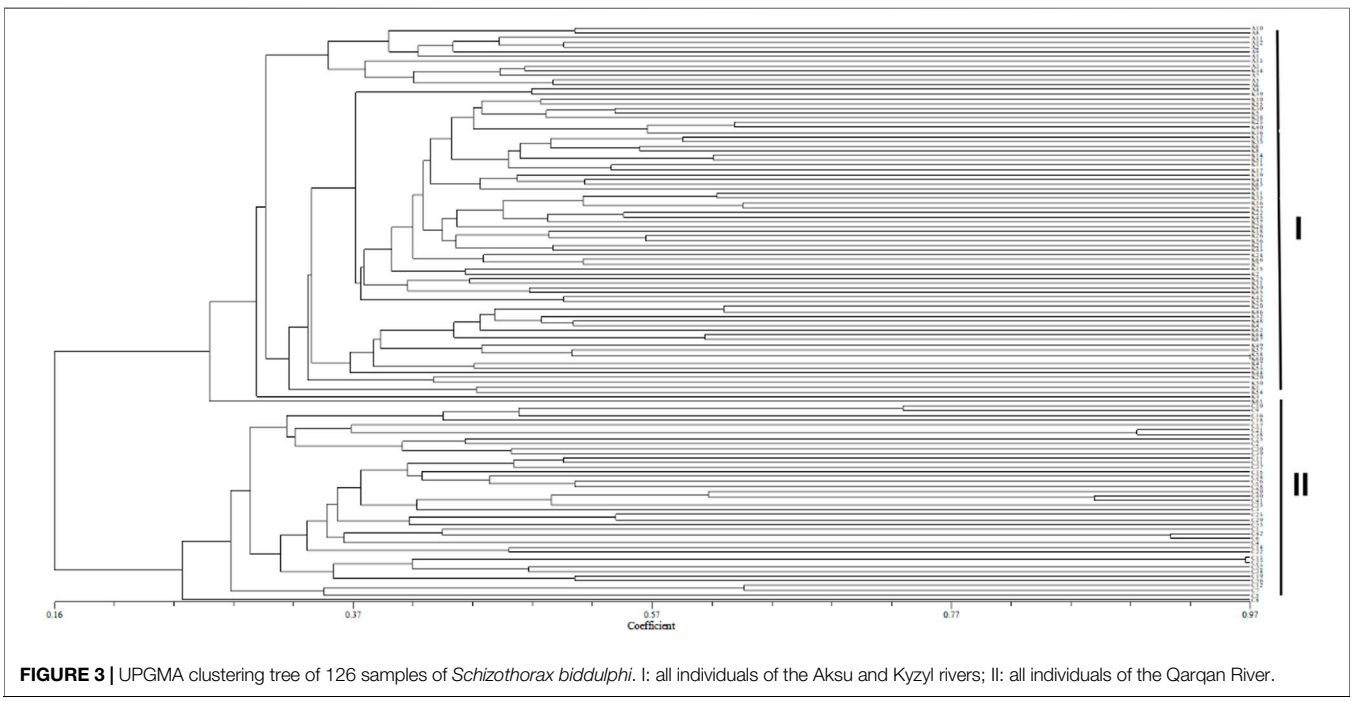


FIGURE 3 | UPGMA clustering tree of 126 samples of *Schizothorax biddulphi*. I: all individuals of the Aksu and Kyzyl rivers; II: all individuals of the Qarqan River.

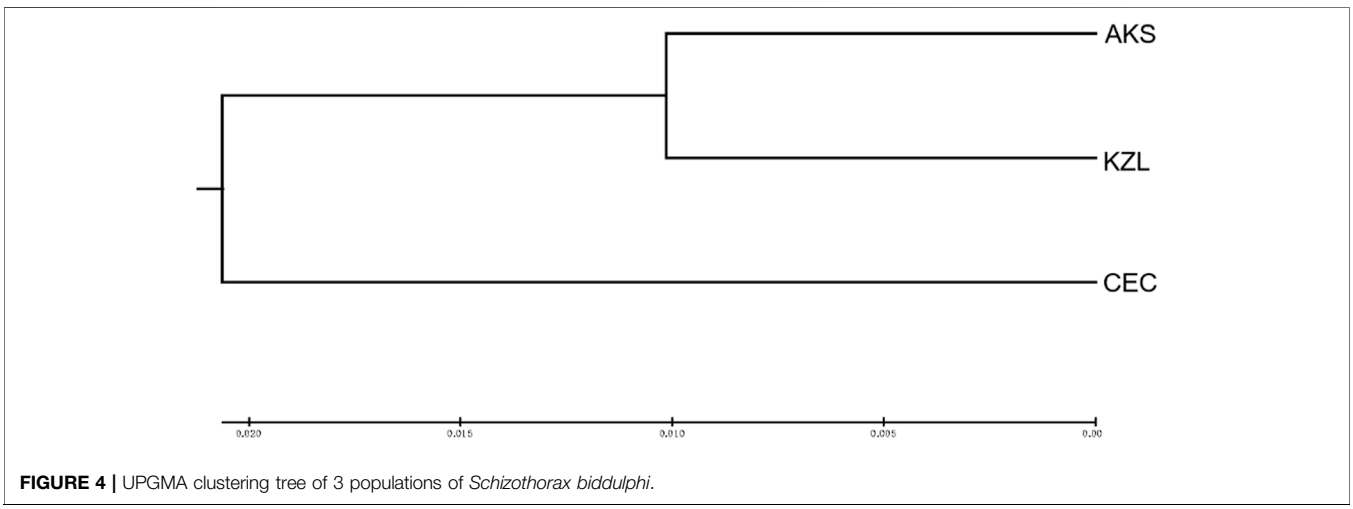


FIGURE 4 | UPGMA clustering tree of 3 populations of *Schizothorax biddulphi*.

TABLE 7 | Analysis of molecular variance (AMOVA) in three populations of *Schizothorax biddulphi*.

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation
Among populations	2	408.829	5.282	20
Within populations	123	2524.794	21.217	80
Total	125	2933.623	26.499	100

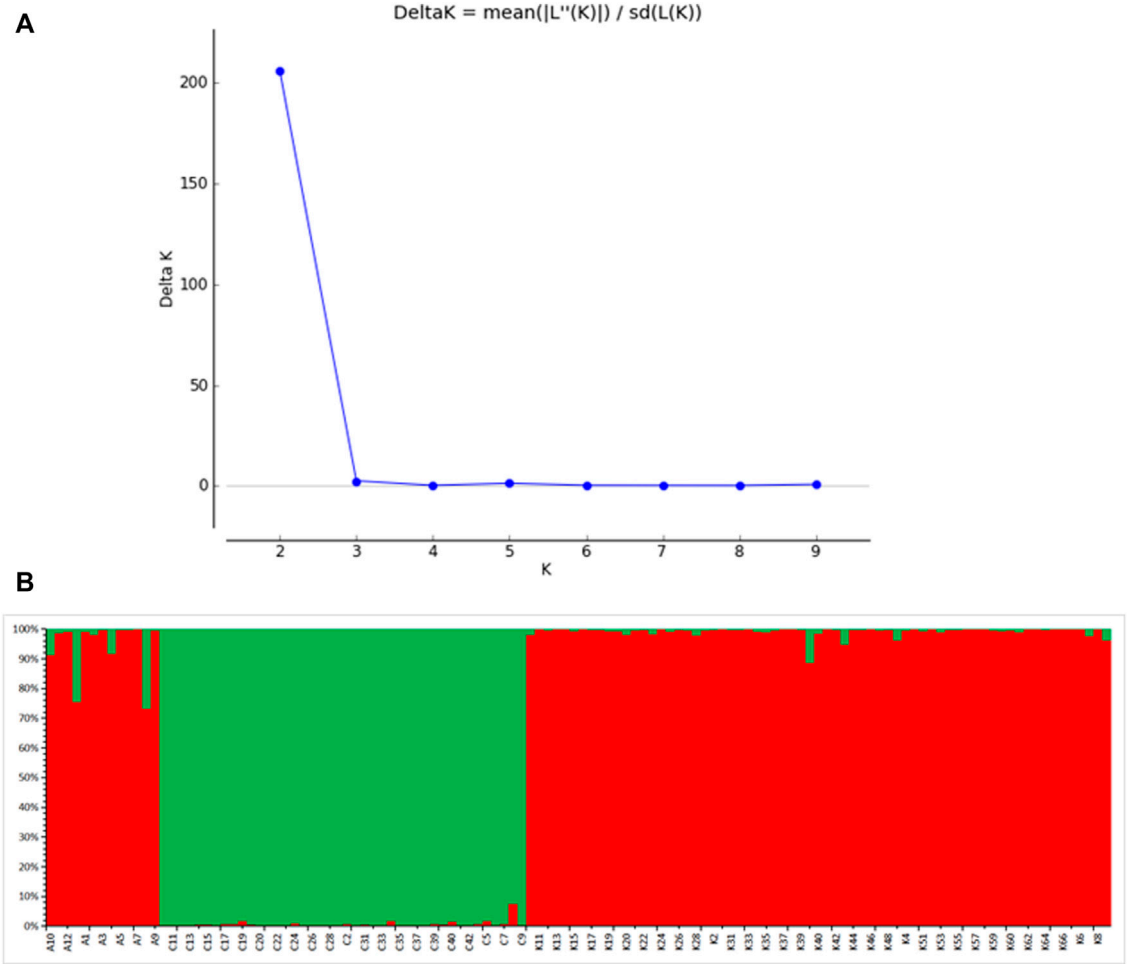


FIGURE 5 | Model choice criterion lnP (D) of the structure analysis for each K value **(A)**; *K* = 2, structure analysis in all three *Schizothorax biddulphi* populations **(B)**, CEC (green), and other two populations (red).

wild populations (**Figure 5B**). When *K* = 2, the gene composition of AKS and KZL samples mainly came from gene bank 1. The gene composition of CEC was mainly derived from gene bank 2 (**Table 8**).

DISCUSSION

SSR Locus Development of *Schizothorax biddulphi*

S. biddulphi is an economic fish only distributed in the Tarim River system of Xinjiang, China, and is endangered at present. However, there are few nucleotide sequences related to *S. biddulphi* microsatellites found in the GeneBank database, and the number of markers that can be used for population genetic diversity research is limited. Therefore, this study obtains microsatellite sequences through genome sequencing of *S. biddulphi*, and develops and screens out microsatellite markers with high polymorphism. SSRs in the genome of *S. biddulphi* are not only abundant but also more numerous, with a predominance of dinucleotide repeats. This is a common phenomenon in the most aquatic animal genome, such as *Takifugu rubripes* (Cui et al., 2006), *Scophthalmus maximus* (Ruan et al., 2011), *Ietalurus*

TABLE 8 | Proportion of ancestry of each population in three gene pools defined with the model-based clustering method.

K = 2	AKS	CEC	KZL
Gene bank 1	0.9415	0.0067	0.9909
Gene bank 2	0.0585	0.9933	0.0091

punetaus (Serapion et al., 2004), *Fugu rubripes* (Edwards et al., 1998), and *Marsupenaeus japonicus* (Lu et al., 2017).

Genetic Diversity of *Schizothorax biddulphi*

Genetic diversity parameters, such as the effective number of alleles, Nei's genetic distance, and Shannon's Information index, can be used to reflect the size of the genetic diversity; in a fixed parameter range, the greater the value of these parameters, the higher the gene richness of the population (Lu et al., 2017). In this study, the average number of alleles of 20 pairs of polymorphic primers in 3 populations of *S. biddulphi* was 19, which was all higher than 9.5, 4.9, and 3.5 of the studies of Gong et al. (2012), Yang et al. (2014), and Luo et al. (2012). However, there was a large difference between the number of alleles at the SSR loci and the number of effective alleles, which indicated that alleles were unevenly distributed in the populations (Kimura and Ohta, 1969). Compared with the polymorphic locus frequency of other related species such as *Schizopygopsis malacanthus* (68.22%), *Schizopygopsis malacanthus chengi* (66.36%), *Schizothorax wangchiachii* (61.21%), *Schizothorax griseus* (61.24%), *Schizothorax grahami* (60.63%), *Schizothorax dolichonema* (57.01%), and *Schizopygopsis stoliczkae* (57.48%), the genetic diversity of *S. biddulphi* is relatively lower (Meng and Zhang, 2007; He et al., 2012). Molecular variance analysis of genetic differences between populations showed that genetic variation was mainly from within populations, with small differences between populations, which was consistent with the result of the study by Yang et al. on Tashikuergan, Duolang headworks, Muzhati River, Yulongkashi River, and Kalakashi River populations (Yang et al., 2014). At the same time, the results of Nei's genetic distance and Shannon's Information index in this study were consistent, indicating that the genetic diversity of CEC was the highest among the three sampling sites, while that of AKS was the lowest. In a finite population, the number of generations required for accidental new selectively neutral mutations depends on the effective population size, and low-density populations favor the accumulation of selectively neutral mutations (Kimura, 1979; Barros et al., 2020). Therefore, the uneven distribution of alleles in the population of *S. biddulphi* may be caused by the decrease in population size. The dams built on the Tarim River in recent years have blocked the migration of fish, leading to a significant increase in the frequency of their inbreeding, which in turn causes low heterozygosity and genetic diversity of the offspring. If this situation is not effectively improved, the genetic diversity of the population will continue to decline (or the inbreeding coefficient will continue to increase), eventually resulting in an irreversible decline of the population and risk of extinction. The genetic variation obtained by microsatellite markers was similar among different populations of *S. biddulphi*, and the genetic variation mainly occurred within populations rather than between populations. CEC may retain the original population resources, or the result may be caused by the small number of AKS samples.

Genetic Structure and Relationship of *Schizothorax biddulphi*

F_{st} and N_m are two important indicators to measure population genetic structure (Pierson et al., 2015). In this study, the genetic differentiation between AKS and KZL was moderate, while that between other populations in pairs was high. However, the size of gene flow between populations was the opposite; that is, there was more gene exchange between AKS and KZL than between other populations. This may be due to the Kizil River being an upstream tributary of the Kashgar River, which originates from the eastern parts of the Pamir Mountains (Hayward, 1869). The Kashgar River eventually flows eastward into the Tarim River, while the Aksu River is the main tributary of the Tarim River, so there is a wide range of gene exchange between the two populations (Wang et al., 2021). Although the Kashgar River has been separated from the mainstream of the Tarim River at present, the separation time is relatively short, and the resulting genetic differentiation is relatively small. The reason for the large genetic differentiation between CEC and the other two populations is as follows: As early as 220 BC, the Tarim River and Cherchen River converged in Lop Nur (Xinjiang, China), and the gene exchange between aquatic organisms in the two rivers was directly influenced by the evolution of Lop Nur (Haysa et al., 2016). Studies have shown that Lop Nur appeared in the Upper Miocene about 5 mega annum ago and dried up at the end of the Pliocene due to climatic reasons. Subsequently, the climate gradually improved and the dryness decreased during 2.7–1.5 mega annum ago, and the surface of Lop Nur recovered before 1.5 mega annum (Yuan and Yuan, 1998; Luo et al., 2006; Chang and Chang, 2013). The genetic differentiation between CEC and other populations may have been caused during the period of Lop Nur drying, which caused geographical isolation between CEC and other populations and the fragmentation of its habitat. Although Cherchen River and Tarim River intersect at Taitema Lake, the genetic differentiation between CEC and other populations still exists significantly. For aquatic organisms, geographical isolation barriers between different water areas usually lead to obvious population genetic differentiation, so the pattern of fish distribution often determines its genetic differentiation pattern (Bergek and Björklund, 2007; Riginos et al., 2011). For example, the Arctic grayling (*Thymallus arcticus*) in the Itkillik, Kuparuk, and Sagavanirktok basins in the foothills of the Brooks Mountains undergo altered dispersal and gene flow due to movement restrictions caused by river drying (Golden et al., 2021).

Based on Nei's genetic distance, the UPMGA method was used for cluster analysis of three *S. biddulphi* populations and individuals. The results showed that AKS and KZL clustered into a cluster, and CEC clustered into a cluster. Structure cluster analysis also shows that the CEC population had great genetic differences from the other two populations, which further verifies the Cherchen River due to early break with the Tarim River, cutoff of the genes within or between is a kind of communication, increase the genetic distance, and then the genetic similarity between CEC and the other two populations was small. After

1950s, although the Kizil River also gradually lost its connection with the Tarim River, the time of geographical isolation between KZL and AKS was relatively short, so the genetic distance between the two populations was relatively close and the genetic similarity coefficient was high.

CONCLUSION

In this study, 20 new SSR markers of *S. biddulphi* were developed by genome survey. These markers used were informative enough and could detect genetic diversity among *S. biddulphi*. The results of this study indicated that Cherchen River had preserved relatively primitive *S. biddulphi* population resources due to its early separation from the mainstream of Tarim River, and showed higher genetic diversity than the other two populations. However, the genetic diversity of *S. biddulphi* was still relatively low compared with other schizothoracin. If it is not its protection and intervention, *S. biddulphi* gene flow between populations appears greatly reduced or even cutoff and will lead to further genetic differentiation between populations, and make it into a vicious circle. Therefore, we must protect the genetic resources of *S. biddulphi* scientifically and rationally, strengthen the protection of its ecological water environment, and improve gene exchange between populations (Caballero and Toro, 2002, Jasim Aljumaili et al., 2018, Roques et al., 2016).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA818042.

ETHICS STATEMENT

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Scientific Ethics Committee of Tarim University.

AUTHOR CONTRIBUTIONS

ZN and JW: conceptualization. YR and LZ: experimental operation. YR and RG: field sampling. ZN and YR: sample determination. ZN: writing—original draft preparation. JW: writing—reviewing, and editing. All authors read and approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (31860729 and 31560721); The Young and Middle-aged Science and Technology Innovation Leading Talent Program Pro-ject of Xinjiang Production and Construction Corps (2018CB033); United Fund of the Ocean University of China and Tarim University (ZHYLH201902); President's Fund of Tarim University-Germplasm resources and Genetic Breeding of Schizothorax in Southern XinjiangInnovative research team project (TDZKCX202204).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.908367/full#supplementary-material

REFERENCES

Barros, J., Winkler, F. M., and Velasco, L. A. (2020). Assessing the Genetic Diversity in Argopecten Nucleus (Bivalvia: Pectinidae), a Functional Hermaphrodite Species with Extremely Low Population Density and Self-fertilization: Effect of Null Alleles. *Ecology Evol.* 10, 3919–3931. doi:10.1002/ece3.6080

Bergek, S., and Björklund, M. (2007). Cryptic Barriers to Dispersal within a Lake Allow Genetic Differentiation of Eurasian Perch. *Evolution* 61, 2035–2041. doi:10.1111/j.1558-5646.2007.00163.x

Caballero*, A., and Toro, M. A. (2002). Analysis of Genetic Diversity for the Management of Conserved Subdivided Populations. *Conserv. Genet.* 3, 289–299. doi:10.1023/a:1019956205473

Chang, Q., and Chang, H. (2013). The Environmental Magnetism Study of Core Ls2 in Lop Nur, Tarim since 7.1Ma. *Quat. Sci.* 33, 876–888. doi:10.3969/j.issn.1001-7410.2013.05.06

Cui, J., Shen, X., Yang, G., Gong, Q., and Gu, Q. (2006). The Analysis of Simple Sequence Repeats in *Takifugu rubripes* Genome. *Period. Ocean. Univ. China.* 2, 249–254. doi:10.16441/j.cnki.hdxb.2006.02.015

Edwards, Y. J. K., Elgar, G., Clark, M. S., and Bishop, M. J. (1998). The Identification and Characterization of Microsatellites in the Compact Genome of the Japanese Pufferfish, *Fugu Rubripes*: Perspectives in Functional and Comparative Genomic Analyses. *J. Mol. Biology* 278, 843–854. doi:10.1006/jmbi.1998.1752

Elmeer, K., Matatt, I., Al-Malki, A., and Alam, M. S. (2019). Genetic Diversity and Relatedness in Arabian oryx (*Oryx leucoryx*) Revealed by SSR Markers. *QScience Connect.* 2019, 3. doi:10.5339/connect.2019.3

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the Number of Clusters of Individuals Using the Software Structure: a Simulation Study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294x.2005.02553.x

Golden, H. E., Holsinger, K. E., Deegan, L. A., Mackenzie, C. J. A., and Urban, M. C. (2021). River Drying Influences Genetic Variation and Population Structure in an Arctic Freshwater Fish. *Conserv. Genet.* 22, 369–382. doi:10.1007/s10592-021-01339-0

Gong, X., Cui, Z., and Wang, C. (2012). Isolation and Characterization of Polymorphic Microsatellite Loci from the Endangered Tarim Schizothoracin (Schizothorax Biddulphi Günther). *Conserv. Genet. Resour.* 4, 795–797. doi:10.1007/s12686-012-9646-1

Guo, X.-Z., Zhang, G.-R., Wei, K.-J., Ji, W., Yan, R.-J., Wei, Q.-W., et al. (2019). Phylogeography of the Threatened Tetraploid Fish, *Schizothorax Waltoni*, in the Yarlung Tsangpo River on the Southern Qinghai-Tibet Plateau: Implications for Conservation. *Sci. Rep.* 9, 2704. doi:10.1038/s41598-019-39128-y

Hayward, G. W. (1869). Journey from Leh to Yarkand and Kashgar, and Exploration of the Sources of the Yarkand River. *Proc. R. Geogr. Soc. Lond.* 14, 41–74. doi:10.2307/1799609

He, Y., An, M., and Li, X. (2012). "RAPD Analysis on Genetic Diversity of Three Kind of Schizothoraxs in Guizhou,". Editor J. Guizhou, 28, 1

Jasim Aljumaili, S., Rafii, M. Y., Latif, M. A., Sakimin, S. Z., Arolu, I. W., Miah, G., et al. (2018). Genetic Diversity of Aromatic Rice Germplasm Revealed by SSR Markers. *Biomed. Res. Int.* 2018, 7658032. doi:10.1155/2018/7658032

Kang, S. W., Patnaik, B. B., Hwang, H.-J., Park, S. Y., Chung, J. M., Song, D. K., et al. (2017). Sequencing and De Novo Assembly of Visceral Mass Transcriptome of the Critically Endangered Land Snail Satsuma Myomphala : Annotation and SSR Discovery. *Comp. Biochem. Physiology Part D Genomics Proteomics* 21, 77–89. doi:10.1016/j.cbd.2016.10.004

Kimura, M. (1979). Model of Effectively Neutral Mutations in Which Selective Constraint Is Incorporated. *Proc. Natl. Acad. Sci. U.S.A.* 76, 3440–3444. doi:10.1073/pnas.76.7.3440

Kimura, M., and Ohta, T. (1969). The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics* 61, 763–771. doi:10.1093/genetics/61.3.763

Le, P., and Chen, Y. (1998). *China Red Data Book of Endangered Animals: Pisces*. Beijing: Science Press.

Li, D., Yao, Y., Huang, X., Cheng, A., Xu, H., Ni, Q., et al. (2012). Simple Sequence Repeat (SSR) Polymorphisms and Population Genetics in Sichuan Wild Rhesus Macaques. *Adv. Mat. Res.* 343-344, 690–697. doi:10.4028/www.scientific.net/AMR.343-344.690

Li, Q., Zheng, X., and Yu, R. (2008). Inheritance Mode of Microsatellite Loci and Their Use for Kinship Analysis in the Pacific Oyster (*Crassostrea gigas*). *Chin. J. Ocean. Limnol.* 26, 256–262. doi:10.1007/s00343-008-0256-4

Liu, Y., Chen, Y., Gong, Q., Lai, J., Du, J., and Deng, X. (2017). Paternity Assignment in the Polyploid *Acipenser Dabryanus* Based on a Novel Microsatellite Marker System. *PLoS One* 12, e0185280. doi:10.1371/journal.pone.0185280

Lu, X., Luan, S., Kong, J., Hu, L., Mao, Y., and Zhong, S. (2017). Genome-wide Mining, Characterization, and Development of Microsatellite Markers in *Marsupenaeus japonicus* by Genome Survey Sequencing. *Chin. J. Ocean. Limnol.* 35, 203–214. doi:10.1007/s00343-016-5250-7

Luo, C., Peng, Z., Yang, D., Liu, W., and He, J. (2006). Research on the Environmental Evolution of Lop-Nur in Xinjiang, China. *Chin. J. Nat.* 28, 37

Luo, W., Nie, Z., Zhan, F., Wei, J., Wang, W., and Gao, Z. (2012). Rapid Development of Microsatellite Markers for the Endangered Fish Schizothorax Biddulphi (Günther) Using Next Generation Sequencing and Cross-Species Amplification. *Int. J. Mol. Sci.* 13, 14946–14955. doi:10.3390/ijms131114946

Marasinghe, M. S. L. R. P., Nilanthi, R. M. R., Hathurusinghe, H. A. B. M., Sooriyabandara, M. G. C., Chandrasekara, C. H. W. M. R. B., Jayawardana, K. A. N. C., et al. (2021). Revisiting Traditional SSR Based Methodologies Available for Elephant Genetic Studies. *Sci. Rep.* 11, 8718. doi:10.1038/s41598-021-88034-9

Meng, L., and Zhang, J. (2007). RAPD Analysis of Genetic Relationships Among 5 Species (Subspecies) of Schizothoracinae in the Yalong River. *China Fish.* 6, 73–76. doi:10.3969/j.issn.1002-6681.2007.06.046

Mohanty, P., Sahoo, L., Pillai, B. R., Jayasankar, P., and Das, P. (2016). Genetic Divergence in Indian Populations ofM. Rosenbergiiusing Microsatellite Markers. *Aquac. Res.* 47, 472–481. doi:10.1111/are.12508

Nie, Z.-L., Wei, J., Ma, Z.-H., Zhang, L., Song, W., Wang, W.-M., et al. (2014). Morphological Variations of Schizothoracinae Species in the Muzhati River. *J. Appl. Ichthyol.* 30, 359–365. doi:10.1111/jai.12376

Peakall, R., Smouse, P. E., Takezaki, N., Nei Takezaki, N., Nei, M., and Tamura, K. (20122010). GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-Aan updatePOPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. *BioinformaticsMol. Biol. Evol.* 2827, 2537747–2539752. doi:10.1093/bioinformatics/bts460

Pierson, J., Luikart, G., and Schwartz, M. (2015). "The Application of Genetic Indicators in Wild Populations: Potential and Pitfalls for Genetic Monitoring," in *Indicators and Surrogates of Biodiversity and Environmental Change*. Editors D. B. Lindenmayer, P. Barton, and J. C. Pierson (Melbourne: CSIRO Publishing), 149

Qiu, J., Guo, R., Li, Y., Zhang, Y., Jia, K., Lei, Y., et al. (2021). De Novo transcriptome Assembly, Functional Annotation and SSR Marker Discovery of Qinling Takin (*Budorcas Taxicolor Bedfordi*). *Animals* 11, 2366. doi:10.3390/ani11082366

Riginos, C., Douglas, K. E., Jin, Y., Shanahan, D. F., and Treml, E. A. (2011). Effects of Geography and Life History Traits on Genetic Differentiation in Benthic Marine Fishes. *Ecography* 34, 566–575. doi:10.1111/j.1600-0587.2010.06511.x

Roques, S., Berrebi, P., Chèvre, P., Rochard, E., and Acolas, M. L. (2016). Parentage Assignment in the Critically Endangered European Sturgeon (*Acipenser sturio*) Based on a Novel Microsatellite Multiplex Assay: a Valuable Resource for Restocking, Monitoring and Conservation Programs. *Conserv. Genet. Resour.* 8, 313–322. doi:10.1007/s12686-016-0538-7

Ruan, X., Wang, W., Kong, J., and Hu, J. (2011). Isolation and Analysis of Microsatellites in the Genome of Turbot (*Scophthalmus maximus* L.). *Afr. J. Biotechnol.* 10, 507–518. doi:10.5897/AJB10.482

Serapion, J., Kucuktas, H., Feng, J., and Liu, Z. (2004). Bioinformatic Mining of Type I Microsatellites from Expressed Sequence Tags of Channel Catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* 6, 364–377. doi:10.1007/s10126-003-0039-z

Tautz, D., and Renz, M. (1984). Simple Sequences Are Ubiquitous Repetitive Components of Eukaryotic Genomes. *Nucl. Acids Res.* 12, 4127–4138. doi:10.1093/nar/12.10.4127

Wang, J., and Scribner, K. T. (2014). Parentage and Sibship Inference from Markers in Polyploids. *Mol. Ecol. Resour.* 14, 541–553. doi:10.1111/1755-0998.12210

Wang, X., Gan, X., Li, J., Chen, Y., and He, S. (2016). Cyprininae Phylogeny Revealed Independent Origins of the Tibetan Plateau Endemic Polyploid Cyprinids and Their Diversifications Related to the Neogene Uplift of the Plateau. *Sci. China Life Sci.* 59, 1149–1165. doi:10.1007/s11427-016-0007-7

Wang, X., Luo, Y., Sun, L., and Shafeeque, M. (2021). Different Climate Factors Contributing for Runoff Increases in the High Glacierized Tributaries of Tarim River Basin, China. *J. Hydrology Regional Stud.* 36, 100845. doi:10.1016/j.ejrh.2021.100845

Yang, T., Meng, W., Gao, T., Guo, Y., Ma, Y., and Zhang, F. (2014). Genetic Polymorphism of Microsatellite DNA in *Schizothorax Biddulphi* Populations. *Arid Zone Res.* 31, 1109–1114. doi:10.13866/j.azr.2014.06.19

Yuan, G., and Yuan, L. (1998). An Approach to the Environmental Changes in Lop-Nur History. *Acta Geogr. Sin.* 53, 83–89. doi:10.11821/xb1998s1011

Zhu, B., Zhou, F., Cao, H., Shao, Z., Zhao, N., May, B., et al. (2002). Analysis of Genetic Variation in the Chinese Sturgeon, *Acipenser Sinensis*: Estimating the Contribution of Artificially Produced Larvae in a Wild Population. *J. Appl. Ichthyol.* 18, 301–306. doi:10.1046/j.1439-0426.2002.00379.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 26 July 2022
doi: 10.3389/fgene.2022.889672

Genetic Structure of the Endangered Coral *Cladocora caespitosa* Matches the Main Bioregions of the Mediterranean Sea

Mar Repullés[1†], *Violeta López-Márquez*[1†], *José Templado*[1], *Marco Taviani*[2,3] *and Annie Machordom*[1*]

[1]*Department Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, MNCN (CSIC), Madrid, Spain,* [2]*ISMAR-CNR, Istituto di Scienze Marine, Consiglio Nazionale delle Ricerche, Bologna, Italy,* [3]*Stazione Zoologica Anton Dohrn, Naples, Italy*

OPEN ACCESS

Edited by:
Narongrit Muangmai, Kasetsart University, Thailand

Reviewed by:
Federica Costantini, Geological and Environmental Sciences, University of Bologna, Italy
Axayacatl Rocha-Olivares, Center for Scientific Research and Higher Education in Ensenada (CICESE), Mexico

***Correspondence:**
Annie Machordom
annie@mncn.csic.es

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *04 March 2022*
Accepted: *21 June 2022*
Published: *26 July 2022*

Citation:
Repullés M, López-Márquez V, Templado J, Taviani M and Machordom A (2022) Genetic Structure of the Endangered Coral Cladocora caespitosa *Matches the Main Bioregions of the Mediterranean Sea. Front. Genet. 13:889672. doi: 10.3389/fgene.2022.889672*

Population connectivity studies are a useful tool for species management and conservation planning, particular of highly threatened or endangered species. Here, we evaluated the genetic structure and connectivity pattern of the endangered coral *Cladocora caespitosa* across its entire distribution range in the Mediterranean Sea. Additionally, we examined the relative importance of sexual and asexual reproduction in the studied populations and their genetic diversity. A total of 541 individuals from 20 localities were sampled and analysed with 19 polymorphic microsatellite markers. Of the genotyped individuals, 482 (89%) had unique multilocus genotypes. Clonality percentages of the populations varied from 0% (in eight populations) to nearly 69% (in one population from Crete). A heterozygosity deficit and a high degree of inbreeding was the general trend in our data set. Population differentiation in *C. caespitosa* was characterised by significant pairwise F_{ST} values with lower ones observed at an intraregional scale and higher ones, between populations from different biogeographic regions. Genetic structure analyses showed that the populations are divided according to the three main sub-basins of the Mediterranean Sea: the Western (Balearic, Ligurian and Tyrrhenian seas), the Central (Adriatic and Ionian seas) and the Eastern (Levantine and Aegean seas), coinciding with previously described gene flow barriers. However, the three easternmost populations were also clearly separated from one another, and a substructure was observed for the other studied areas. An isolation-by-distance pattern was found among, but not within, the three main population groups. This substructure is mediated mainly by dispersal along the coastline and some resistance to larval movement through the open sea. Despite the low dispersal ability and high self-recruitment rate of *C. caespitosa*, casual dispersive events between regions seem to be enough to maintain the species' considerable genetic diversity. Understanding the population connectivity and structure of this endangered scleractinian coral allows for more informed conservation decision making.

Keywords: asexual reproduction, *Cladocora caespitosa*, clones, dispersion, marine connectivity, microsatellites, population structure

INTRODUCTION

Marine population dynamics is determined by connectivity among populations, which can play a key role in species resilience (Cowen and Sponaugle, 2009). Gene flow between geographically separated populations of marine species is mainly governed by both the biological traits of species and oceanographic dynamics (Villamor et al., 2014). Historical barriers, currents, habitat discontinuities and shoreline configuration, along with the larval life span and behaviour of a species, can all drive genetic differentiation between populations (González-Wangüemert et al., 2010). For instance, in the Mediterranean Sea, several currents and fronts have been described as potential barriers to gene flow (Coll et al., 2010; Villamor et al., 2014; Pascual et al., 2017), which could influence the genetic structure of species. Larval dispersal capacity, settlement success and survival of newly settled post-metamorphic individuals until recruitment are also crucial factors affecting species distribution and population connectivity (Botsford et al., 2001; Pascual et al., 2017). Population genetics studies are essential to estimate the degree of connectivity and to identify associated processes that may influence it, such as barriers to gene flow, self-recruitment and population isolation or fragmentation, among others.

In this study, we examine the genetic connectivity of populations of *Cladocora caespitosa* (Linnaeus, 1767), a Mediterranean colonial and zooxanthellate scleractinian coral whose features are similar to typical tropical reef-building corals. This species is, for instance, able to form extensive bioherms that may fuse in reef-like structures (Kružić et al., 2008). Records show that *Cladocora* has been present in the Mediterranean basin since the Miocene (Vertino et al., 2014). Although morphologically indistinguishable from other *Cladocora* fossils, *C. caespitosa* is thought to have been present in the basin since, at least, the warm-temperate late Pliocene (Aguirre and Jiménez, 1998; Peirano et al., 2009; Bosio et al., 2021), and was particularly common in the late Pleistocene (Cuerda et al., 1986; Bernasconi et al., 1997; Amorosi et al., 2014). The species is currently distributed throughout the entire Mediterranean basin as discontinuous and isolated colonies or, more rarely, as coral beds or banks (Schiller, 1993; Peirano et al., 2004; Kersting and Linares, 2012; Chefaoui et al., 2017; López-Márquez et al., 2019, 2021). Though numerous studies on the ecology and biological traits of *C. caespitosa* have been carried out (Peirano et al., 2004; Kersting et al., 2013a, 2013b, 2014a, 2014b; Rubio-Portillo et al., 2018; Pons-Fita et al., 2020, among others), genetic connectivity studies of this species are scarce. Only three studies on the genetic differentiation of the species, all conducted at a regional scale, have been published: Casado-Amezúa et al. (2014) studied the species in the western Mediterranean where they found low genetic connectivity related to sporadic dispersal events among the studied populations, almost the same results obtained in the eastern Mediterranean by López-Márquez et al. (2021). In the third study, conducted in the Adriatic Sea and adjacent Ionian, López-Márquez et al. (2019) found that the connectivity patterns were mainly driven by the shoreline configuration.

Cladocora caespitosa is found in a wide variety of environments, from shallow waters to about 35 m of depth (Bellan-Santini et al., 2007). It can resist strong currents but is sensitive to high wave impact (Chefaoui et al., 2017). This emblematic species displays both sexual and asexual reproduction. Asexual reproduction can occur by fragmentation or polyp removal (Kružić et al., 2008) or by asexual buds produced by polyps (Rodolfo-Metalpa et al., 2008) that generate individuals with identical multilocus genotypes ("ramets"), which, in turn, form "genets" of potentially different sizes (Baums, 2008). Sexual reproduction in *C. caespitosa* is generally seasonal and synchronous; however, differences have been observed in distinct geographic areas. In the western and eastern Mediterranean, gonochoric colonies release gametes at the end of summer, when temperatures begin to fall (Kersting et al., 2013b; Hadjioannou, 2019). By contrast, in the Adriatic Sea, hermaphroditic colonies reproduce at the beginning of the summer, when temperatures begin to rise, typically coinciding with a full moon (Kružić et al., 2008). It is unknown if these reproductive differences are associated with genetic differences.

Cladocora caespitosa might have been a keystone species of Mediterranean benthic communities in the past, thus playing an important role in the biodiversity of this marine realm. In fact, bioconstructions formed by this coral are known to harbour a high diversity of micro- and macrofauna (Koukouras et al., 1998; Pitacco et al., 2014). Unfortunately, like many others Mediterranean marine species (Bianchi and Morri, 2000), *C. caespitosa* is currently in alarming decline (Kersting et al., 2014b) and has been categorised as an endangered species in the IUCN Red List (Casado de Amezua et al., 2015).

The main goal of the present study is to evaluate the genetic structure and connectivity pattern of *C. caespitosa* across its entire distribution range, which has not been done to date. This will allow us to have a more complete understanding of the level of connectivity and differentiation among populations, and the potential conservation implications they may have for the species. Although local populations of marine species are generally considered demographically open (Ward et al., 1994; Hellberg, 2007), high connectivity among populations of this coral would not be expected, given that eggs are negatively buoyant (Kružić et al., 2008), limiting their planktonic dispersal and favouring local retention (Kersting et al., 2014a).

We hypothesise that *C. caespitosa* has a strong population structure with genetic differentiation between populations. We also assess the relative importance of sexual versus asexual reproduction in the studied populations, and its impact on genetic diversity as some of the populations have experienced adverse conditions that tend to increase asexual reproduction, which could influence population structure (López-Márquez et al., 2021). We hypothesise that the clonal structure of *C. caespitosa* varies across its geographic range due to the influence of various extrinsic factors affecting sexual and asexual reproductive potential.

Given the status of *C. caespitosa* as an endangered species, efficient conservation management is necessary for its long-term survival. Population connectivity studies are a useful tool for this

TABLE 1 | List of the sampling localities of *C. caespitosa.* For each location, sub-basin and country are indicated in parentheses. Also provided are the population codes, GPS coordinates and number of samples (*N*) collected from each locality.

Location	Code	GPS Coordinates	*N*
Cabo de Palos, Murcia (Western, Spain)	CDP	37°37′42.90"N 0°42′7.33"O	30
Punta Gavina, Formentera (Western, Spain)	GAV	38°43′6.06"N 1°22′46.68"E	20
Isla Espardell, Formentera (Western, Spain)	ESP	38°47′16.55"N 1°28′13.02"E	31
Puerto Tofiño, Columbretes (Western, Spain)	COL	39°57′10.93"N 0°41′47.08"E	31
L'Amtella, Tarragona (Western, Spain)	PUN	40°50′26.25"N 0°44′58.92"E	30
Na Macaret, Menorca (Balearic, Spain)	MEN	40° 0′58.09"N 4°12′10.21"E	21
Palau (Tyrrhenian, Italy)	PAL	41°11"15.40"N 9°23"2.99"E	18
Bonassola (Ligurian, Italy)	BON	44°10′50.42"N 9°34′53.71"E	20
Framura (Ligurian, Italy)	BON	44°12′2.79"N 9°33′8.95"E	9
Porto Cesareo (Ionian, Italy)	POC	40°11′ 715″N 17° 55′ 077″E	35
San Foca, Otranto (Adriatic, Italy)	OTR	40° 06′ 554″N 18° 31′ 153″E	35
Torre Guaceto (Adriatic, Italy)	TOG	40° 42′ 999″N 17° 48′ 003″E	35
Tremiti Island (Adriatic, Italy)	TRE	42° 8.315′ N 15° 31.437′ E	35
Porec (Adriatic, Croatia)	POR	45°13′53.15"N 13°35′16.36"E	14
Kornati (Adriatic, Croatia)	KOR	43° 916′ 118″N 15° 146′ 881″E	35
Boka Kotorska (Adriatic, Montenegro)	BOK	42° 23′ 252″N 18° 34′ 178″E	34
Crete (Cretan, Greece)	CRE	35° 1′46.38"N 24°39′2.00"E	16
Nea Peramos (Aegean, Greece)	NEA	40° 49′31.9"N 24°20′01.9"E	31
Liopetri (Levantine, Cyprus)	LIO	34° 57′30.2"N 33°54′05.7"E	31
Kryo Nero (Levantine, Cyprus)	KRY	34° 58′57.0"N 34°01′00.8"E	30

purpose as they provide data on the resilience and sustainability of a species (Gaeta et al., 2020). By providing an overview of the connectivity of an emblematic species of the Mediterranean Sea, a key factor that must be considered for conservation plans and for decision-making on the design of marine protected areas (Holland et al., 2017), we hope to contribute to the persistence of this coral species.

MATERIALS AND METHODS

Study Site and Sample Collection

A total of 541 individuals from 20 Mediterranean localities were collected by SCUBA diving. Due to the close proximity of Bonassola and Framura (both in the Ligurian Sea), and the relatively low number of specimens sampled there, the two localities were analysed as a single population (BON). Four localities were sampled in the eastern Mediterranean: one each in Crete and Greece and two in Cyprus. Seven localities were sampled in the central Mediterranean in the Ionian Sea or along both coasts of the Adriatic Sea in Italy, Croatia and Montenegro. Nine localities were sampled in the western Mediterranean, in Spain and western Italy (**Table 1**; **Figure 1**).

Some of these localities were previously analysed (Casado-Amezúa et al., 2014; López-Márquez et al., 2019, 2021), however, with two different sets of primers (the first set consisted of eight primer pairs and was used by Casado-Amezúa et al., 2011 and López-Márquez et al., 2019; the second set, used by López-Márquez et al., 2021, consisted of 11 new pairs). López-Márquez et al. (2021) previously analysed three of the four eastern Mediterranean populations (NEA, KRY and LIO) using both primer sets. The results of these previous studies served as a reference and were included here in order to provide a complete picture of the species distribution across the entire Mediterranean. To facilitate comparisons, several specimens representing the analysed populations were also genotyped to ensure the uniform allele assignation with both primer sets. As the Adriatic (OTR, TOG, TRE, KOR and BOK) and Ionian (POC) populations were already analysed using the first primer set (López-Márquez et al., 2019), here we only analysed them using the second set. The newly sampled populations from the western Mediterranean (GAV, ESP, MEN, BON and PAL), the Adriatic (POR) and the eastern Mediterranean (CRE), together with the populations previously studied by Casado-Amezúa et al. (2014) (CDP, CON and PUN), were genotyped using both primer sets.

In order to avoid sampling clonal individuals, colonies collected by SCUBA diving were separated by at least 1 m. Some polyps from each colony were carefully excised and stored in vials with absolute ethanol and preserved at 4°C before processing. All necessary permits were obtained for the field studies.

DNA Extraction and Microsatellite Amplification

The QIAGEN Biosprint 15 DNA Blood Kit (Qiagen) was used for DNA extraction and purification, following the manufacturer's protocol. Each DNA sample was diluted to a final concentration of 0.3 ng/μl.

The DNA samples were genotyped with the 19 polymorphic microsatellite loci previously isolated for *C. caespitosa* [the eight by Casado-Amezúa et al. (2011) and the 11 by López-Márquez et al. (2021)]. The 19 primers pairs were combined in five multiplex reactions at a concentration of between 0.2 and 0.4 μM, and mixed with 1x Qiagen Multiplex PCR Master Mix, 0.3 ng of DNA and water to a total volume of 7 μl. To facilitate the genotyping, the forward primer from each pair was

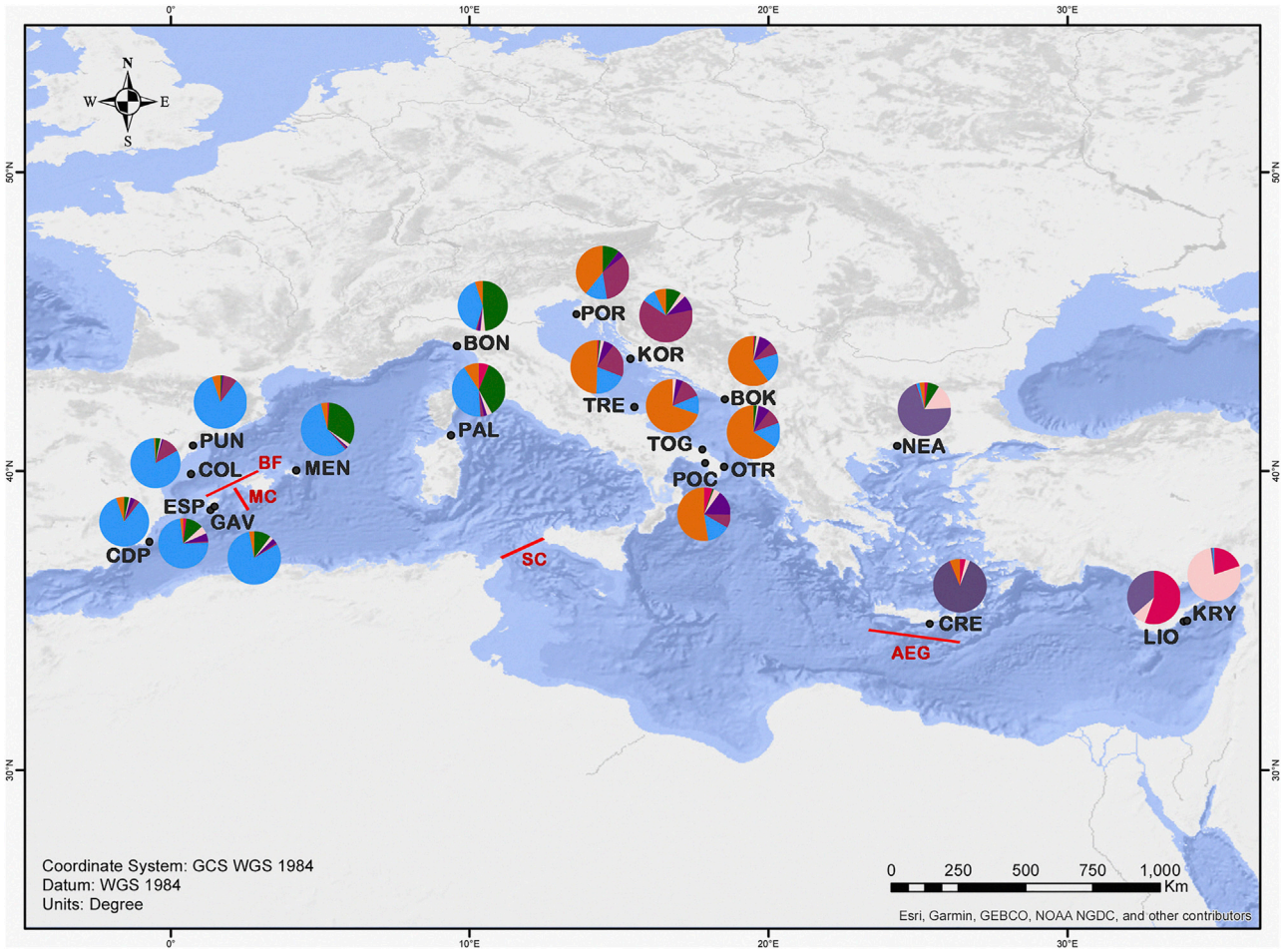


FIGURE 1 | Map showing the locations of the sampled populations of *C. caespitosa* in the Mediterranean Sea. The main barriers detected in our analyses are indicated in red: MC, Mallorca Channel; BF; Balearic Front; SC, Sicily Channel and AEG, Aegeant Front (Ruiz et al., 2009; Pascual et al., 2017). Pie diagrams show a summary of the STRUCTURE results for $K = 7$ (colours as in **Figure 5**), in terms of the proportion of the identified genetic clusters assigned to each site.

end-labelled with either NED, VIC, PET or 6-FAM, and the reverse primers were pig-tailed with 5′-GTTTCTT-3' (López-Márquez et al., 2021). The PCR cycling profile included an initial denaturation step at 95°C for 5 min, followed by 35 cycles at 94°C for 30 s, 56°C for 90 s and 72°C for 30 s, and a final extension step at 72°C for 10 min.

GENEMAPPER software v4.0 (Applied Biosystems) was used to analyse the electropherograms. The presence of null alleles was checked with MICRO-CHECKER v.2.2.3 (Van Oosterhout et al., 2004).

Genotype Analyses and Clonal Structure Parameters

By performing the option "multilocus matches" in GenAlEx 6.5 (Peakall and Smouse, 2006), we calculated the number of both unique (N_g) and non-unique (identical) multilocus genotypes per site, which allowed us to determine the number of clones present in our data set. To analyse clonal structure parameters, and to calculate genotypic richness (standardised), genotypic evenness and genotypic diversity, we followed the methodology of Aranceta-Garza et al. (2012). Also, we classified populations according to the ratio of asexual to sexual reproduction they displayed, which was based on both genotypic evenness and genotypic diversity as in Baums et al. (2006) who delimited four groups: sexual, mostly sexual, mostly asexual and asexual. We also calculated the D index (Baums et al., 2006) as:

$$D = 1 - \left(\frac{\sum n_i (n_i - 1)}{N(N-1)} \right)$$

Where n_i is the number of individuals of genotype i, and N the total number of individuals in the population. D values range between zero and one. If D equals zero, the entire population belong to the same unit of clonal growth.

Genetic Variability

After excluding the individuals identified as clones from the data set, we used GenAlEx 6.5 and GENEPOP v4.0 (Raymond and Rousset, 1995) to calculate allelic diversity (N_a), observed (H_o)

and expected (H_e) heterozygosity and the index F_{IS}, which is commonly used as an inbreeding coefficient, and test for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD). We corrected the significance of p values with the sequential Bonferroni method (Rice, 1989). Prior to performing further analyses with the data set, we look for the presence of loci under selection using two methodologies: a neutrality test performed in ARLEQUIN v3.5 (Excoffier et al., 2005) and a Bayesian approach using BAYESCAN (Foll, 2012). Finally, we estimated effective population size (N_e) with NE ESTIMATOR 2 (Do et al., 2014), following the linkage disequilibrium method (Waples and Do, 2008).

Population Differentiation and Migration in *C. caespitosa*

To assess population structure and differentiation in *C. caespitosa*, we used various methods, including Wright's fixation index (F_{ST}), principal coordinates analysis (PCoA), isolation by distance (IBD), BARRIER analyses, the analysis of molecular variance (AMOVA), STRUCTURE and discriminant analysis of principal components (DAPC).

Population differentiation was estimated with F_{ST} between pairwise sampling sites through Weir and Cockerham's estimators in GENETIX v.4.05 (Belkhir et al., 2004). The negative values obtained because of mathematical artefacts (F_{ST} cannot be negative) were set to zero. Standardised F_{ST} values (F'_{ST}) were calculated in GenAlEx, assuming that each population has different alleles for each locus, which allowed us to estimate the maximum possible value of differentiation (Hedrick, 2005). To visualise a possible pattern of genetic structure, a PCoA was performed in GenAlEx with the obtained F_{ST} values.

To determine whether the genetic structure was driven by population distribution and geographic distance, we quantified IBD as the correlation between linearised F_{ST} (F_{ST}/1- F_{ST}) and log-transformed geographic distance in kilometres using a Mantel test (Mantel, 1967) with 10,000 permutations in GenAlEx. Geographic distance was calculated as the shortest path between sampling locations across the sea. These analyses were also performed independently for the populations in the three Mediterranean sub-basins (western and eastern Mediterranean and the Adriatic Sea). Furthermore, to identify potential barriers to gene flow for *C. caespitosa*, such as oceanographic fronts or currents, pairwise F_{ST} values and population geographic coordinates were included in the analyses performed in BARRIER v2.2 (Manni et al., 2004). To calculate barrier robustness, we generated 100 resampled bootstrap matrices in R (R Development Core Team, 2017), using an R script provided by Eric Petit (UMR ECOBIO CNRS, Paimpont, France).

Population genetic structure was assessed by a Bayesian clustering approach performed in STRUCTURE 2.3.4 (Pritchard et al., 2000). This program calculates population allele frequencies and, on the basis of HWE estimates, the probability of an individual belonging to one of the obtained genetic clusters. As we assumed that individuals can have ancestors from different locations, an admixture model was implemented with correlated allele frequencies and location as a prior. The analysis was run with a burn-in of 10,000 iterations and 100,000 Markov chain Monte Carlo, with a putative K of up to 25 (six clusters more than the number of sampling sites considered in the analysis) and 20 replicate runs. To infer the number of genetic clusters (best value of K), we applied three methods. We first used STRUCTURE HARVESTER (Earl and vonHoldt, 2012) and CLUMPAK (Kopelman et al., 2015) to evaluate the optimal K value following the method proposed by Evanno et al. (2005) to calculate ΔK. CLUMPAK was also employed to identify the K for which Pr ($K = k$) is the highest using ln (Pr (X/K)). Although the Evanno ΔK method may efficiently capture the uppermost level of structure, it may underestimate the number of groups (K) (Puechmaille, 2016). Therefore, we also estimated the number of clusters with the method of Puechmaille (2016) using StructureSelector (Li and Liu, 2018). We used the CLUMPAK web server to find the best alignment of the STRUCTURE results across the 20 replicates for each K.

The R package *adegenet* v2.1.1 (Jombart, 2008) was used to perform the DAPC. This method divides genetic variability of variance between and within groups and then optimises the variance between groups and minimises it within groups (Jombart, 2012), without considering whether population are in HWE.

To quantify the molecular variance, we ran the AMOVA in ARLEQUIN v3.11 (Excoffier et al., 2005) with 1000 permutations for all the populations and with the groups inferred by the STRUCTURE analyses. Putative first generation migrants were identified with a Bayesian assignment method (Rannala and Mountain, 1997) in GENECLASS2 (Piry et al., 2004).

Finally, we tested for recent reductions in effective population size using the allele frequency data in BOTTLENECK v1.2.02 (Cornuet and Luikart, 1996). Simulations were run under three mutation models: the infinite alleles model (IAM), the stepwise mutation model (SMM) and the two-phase mutation model (TPM) (Di Rienzo et al., 1994), with 10,000 iterations and a descriptor of allele frequency distribution ('mode-shift'). Two statistical tests, the Sign test (Cornuet and Luikart, 1996) and the Wilcoxon sign-rank test (Luikart et al., 1998), were used to test each model.

RESULTS

Of the 19 microsatellites used in this study, one (L2) was discarded due to the ambiguity of its results. All of the analysed loci were polymorphic for all populations except L29, which was monomorphic for three of them (LIO, TOG and MEN), and V46 (for LIO).

Null alleles were detected in the populations analysed. As the presence of null alleles is known to inflate measures of genetic differentiation and cause overestimations of F_{ST} (Chapuis and Estoup, 2007), we repeated the population differentiation analyses correcting for the null alleles. However, no significant differences were found between the pairwise F_{ST} and the corrected F_{ST} values; therefore, we decided not to consider the null alleles correction.

TABLE 2 | Genotypic diversity of the analysed populations of *C. caespitosa* based on 19 microsatellites. *N*, number of polyps (colonies) sampled; N_g, number of unique multilocus genotypes per site; N_g/N, genotypic richness; G_o, observed genotypic diversity; G_o/N_g, genotypic evenness; G_e, expected genotypic diversity(*N*); G_o/G_e, genotypic diversity and the *D* index.

Pop	*N*	N_g	N_g/N	G_o	G_o/N_g	G_e	G_o/G_e	*D*
CDP	30	30	1	30	1	30	1	1
GAV	20	19	0.95	18.18	0.96	20	0.91	0.99
ESP	31	28	0.90	24.64	0.88	31	0.80	0.99
COL	31	31	1	31	1	31	1	1
PUN	30	26	0.87	23.68	0.91	30	0.79	0.99
MEN	21	13	0.62	6.58	0.51	21	0.31	0.89
PAL	18	18	1	18	1	18	1	1
BON	29	27	0.93	25.48	0.94	29	0.88	0.99
POC	35	31	0.89	26.06	0.84	35	0.75	0.99
OTR	35	35	1	35	1	35	1	1
TOG	35	35	1	35	1	35	1	1
TRE	35	35	1	35	1	35	1	1
POR	14	14	1	14	1	14	1	1
KOR	35	34	0.97	33.10	0.97	35	0.95	0.99
BOK	34	33	0.97	32.11	0.97	34	0.94	0.99
CRE	16	5	0.31	4.13	0.82	16	0.26	0.81
NEA	31	31	1	31	1	31	1	1
LIO	31	11	0.35	2.33	0.21	31	0.08	0.58
KRY	30	26	0.87	22.50	0.86	30	0.75	0.98

TABLE 3 | Genetic diversity and effective population size estimates for the 18 analysed populations of *C. caespitosa*. N_a, standardised number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient; N_e, effective population size. *populations that are not in HWE.

Pop	N_a	H_o	H_e	F_{IS}	N_e
CDP	4.672	0.530	0.565	0.051*	319.9
GAV	4.278	0.538	0.563	0.042	233.9
ESP	4.242	0.501	0.563	0.118*	83.2
COL	4.232	0.477	0.533	0.097*	33.5
PUN	4.359	0.508	0.550	0.095*	96.6
MEN	4.056	0.509	0.524	0.043	4.9
PAL	4.561	0.532	0.599	0.118*	300.7
BON	4.697	0.532	0.597	0.161*	Infinite
POC	4.429	0.539	0.568	0.061*	19.8
OTR	4.758	0.507	0.542	0.069*	65.2
TOG	4.535	0.503	0.542	0.069*	Infinite
TRE	4.758	0.526	0.571	0.065*	110.7
POR	4.455	0.516	0.563	0.085*	15.0
KOR	4.616	0.532	0.587	0.088*	Infinite
BOK	4.621	0.506	0.559	0.101*	76.8
NEA	4.859	0.544	0.587	0.072*	Infinite
LIO	3.167	0.587	0.496	–0.195*	0.4
KRY	4.116	0.510	0.563	0.130*	4.9
Mean	4.412	0.522	0.560	0.072	—

Clonal Structure in *C. caespitosa*

Of the 541 individuals genotyped, 482 (89%) had unique multilocus genotypes (N_g) (**Table 2**). A total of 59 ramets belonging to 25 genets was detected. Clonality percentages of the populations varied from 0% (CDP, COL, PAL, OTR, TOG, TRE, POR and NEA) to 68.75% in CRE. Genotypic evenness (G_o/N_g) ranged from 0.21 for LIO to one for CDP, COL, PAL, OTR, TOG, TRE, POR and NEA. The lowest values of genotypic richness (N_g/N) were detected for CRE and LIO. Even though these two populations, which have more clonal individuals than the others, had similar values of genetic richness, the observed genotypic diversity (G_o) of LIO was half that of CRE, reflecting a difference in genet size (LIO had fewer but bigger genets). Genotypic diversity (G_o/G_e) varied from 0.08 for LIO, indicating "mostly asexual" reproduction (also for CRE and MEN), to one for CDP, COL, PAL, OTR, TOG, TRE, POR and NEA, indicating "sexual reproduction". The others populations presented "mostly sexual" reproduction by the same criterion. The population with the lowest *D* index value was LIO, agreeing with the results obtained for genotypic diversity and genotypic richness. Due to the low number of unique multilocus genotypes detected for CRE (N_g = 5), we excluded this population from further analysis.

Genetic Variability in *C. caespitosa*

Linkage disequilibrium (LD) among loci was detected in some populations (in KRY for all loci, and in LIO, KOR, POC, COL, POR, TOG and OTR for some of them). Nevertheless, LD was not observed in any of the loci for all the populations; therefore, all loci were considered independent.

Standardised allelic richness (N_a) across loci for each population ranged from 3.167 for LIO to 4.859 for NEA

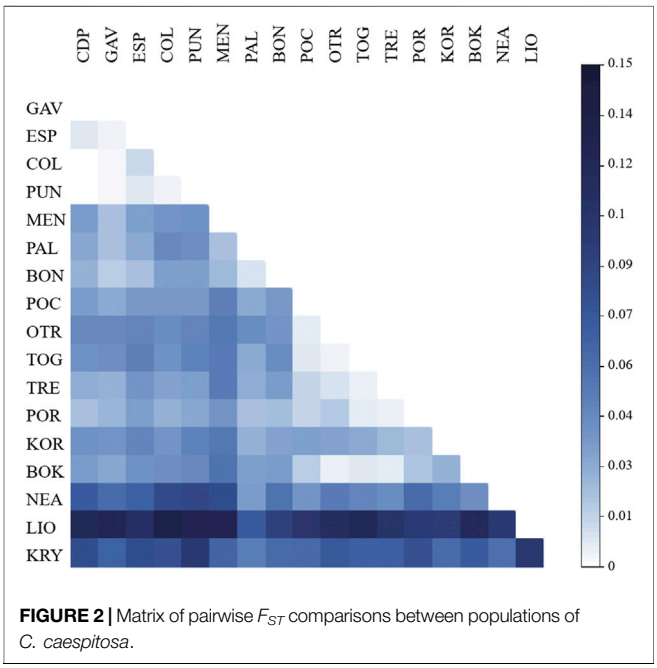


FIGURE 2 | Matrix of pairwise F_{ST} comparisons between populations of *C. caespitosa*.

(**Table 3**). The mean value across all loci and populations was 4.412. Observed heterozygosity (H_o) ranged from 0.477 in COL to 0.587 in LIO. Only LIO presented a H_o value greater than expected (H_e) (0.587 and 0.496, respectively) and, therefore, a negative value for F_{IS} (-0.195), indicating a heterozygote excess. The rest of the populations presented lower values of H_o than H_e and positive F_{IS} values ranging from 0.042 in GAV to 0.16 in BON, indicating a heterozygosity deficit (or a homozygosity excess) and a high degree of inbreeding. All populations,

TABLE 4 | Pairwise F_{ST} values for the analysed populations of *C. caespitosa* (lower diagonal).

	CDP	GAV	ESP	COL	PUN	MEN	PAL	BON	POC	OTR	TOG	TRE	POR	KOR	BOK	NEA	LIO	KRY
CDP	**0**	0	0.019	0	0	0.083	0.075	0.059	0.086	0.103	0.1	0.071	0.051	0.099	0.105	0.174	0.260	0.198
GAV	0	**0**	0.008	0.005	0.005	0.051	0.052	0.038	0.075	0.101	0.103	0.069	0.062	0.096	0.095	0.147	0.263	0.162
ESP	**0.009**	0.0038	**0**	0.031	0.022	0.086	0.082	0.050	0.101	0.112	0.126	0.102	0.086	0.118	0.127	0.170	0.246	0.203
COL	0	0.0028	**0.013**	**0**	0.007	0.087	0.105	0.079	0.087	0.097	0.100	0.084	0.063	0.093	0.118	0.195	0.279	0.183
PUN	0	0.0028	0.009	0.003	**0**	0.095	0.102	0.085	0.094	0.108	0.117	0.090	0.075	0.117	0.133	0.209	0.271	0.236
MEN	**0.037**	**0.029**	**0.036**	**0.04**	**0.042**	**0**	0.054	0.059	0.120	0.121	0.127	0.127	0.096	0.129	0.151	0.197	0.261	0.158
PAL	**0.033**	**0.022**	**0.031**	**0.045**	**0.042**	**0.022**	**0**	0.027	0.080	0.103	0.076	0.074	0.055	0.067	0.098	0.094	0.169	0.134
BON	**0.027**	**0.017**	**0.021**	**0.035**	**0.035**	**0.024**	0.01	**0**	0.096	0.093	0.110	0.093	0.054	0.086	0.113	0.142	0.211	0.158
POC	**0.036**	**0.031**	**0.038**	**0.038**	**0.038**	**0.050**	**0.031**	**0.037**	**0**	0.011	0.027	0.036	0.036	0.080	0.039	0.093	0.231	0.151
OTR	**0.046**	**0.046**	**0.047**	**0.044**	**0.047**	**0.055**	**0.045**	**0.039**	0.006	**0**	0.006	0.018	0.041	0.075	0.016	0.123	0.243	0.172
TOG	**0.041**	**0.043**	**0.05**	**0.041**	**0.048**	**0.053**	**0.031**	**0.044**	**0.007**	0.004	**0**	0.009	0.021	0.073	0.032	0.116	0.264	0.170
TRE	**0.029**	**0.027**	**0.04**	**0.034**	**0.036**	**0.053**	**0.029**	**0.036**	**0.014**	**0.01**	0.005	**0**	0.016	0.059	0.027	0.109	0.239	0.172
POR	**0.022**	**0.027**	**0.035**	**0.028**	**0.032**	**0.04**	**0.022**	**0.023**	**0.014**	**0.02**	0.007	0.005	**0**	0.052	0.065	0.147	0.218	0.198
KOR	**0.042**	**0.04**	**0.047**	**0.04**	**0.048**	**0.054**	**0.028**	**0.033**	**0.034**	**0.034**	**0.03**	**0.024**	**0.022**	**0**	0.071	0.124	0.217	0.149
BOK	**0.037**	**0.032**	**0.041**	**0.042**	**0.045**	**0.058**	**0.036**	**0.037**	**0.017**	0.005	**0.008**	**0.007**	**0.02**	**0.028**	**0**	0.111	0.268	0.175
NEA	**0.074**	**0.062**	**0.068**	**0.084**	**0.088**	**0.083**	**0.037**	**0.058**	**0.039**	**0.053**	**0.047**	**0.043**	**0.062**	**0.051**	**0.042**	**0**	0.228	0.144
LIO	**0.12**	**0.122**	**0.110**	**0.134**	**0.126**	**0.124**	**0.074**	**0.092**	**0.102**	**0.114**	**0.12**	**0.105**	**0.099**	**0.097**	**0.118**	**0.1**	**0**	0.228
KRY	**0.083**	**0.067**	**0.083**	**0.08**	**0.099**	**0.067**	**0.052**	**0.062**	**0.064**	**0.076**	**0.07**	**0.07**	**0.08**	**0.063**	**0.073**	**0.059**	**0.101**	**0**

Significant values are in bold (p < 0.05). F'_{ST} *values are shown in the upper diagonal.*

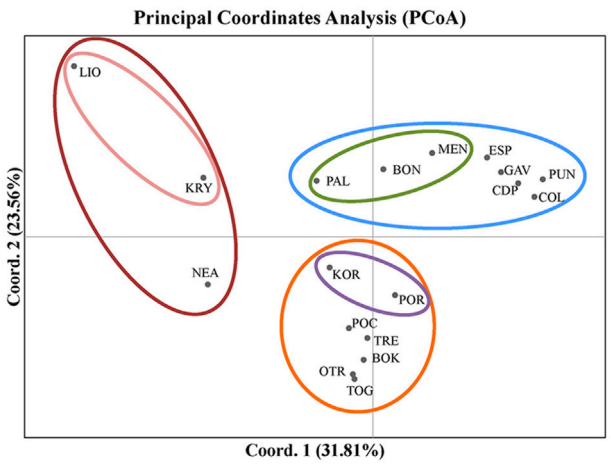


FIGURE 3 | Results of the Principal Coordinates Analysis (PCoA) used to detect clustering of the populations of *C. caespitosa* on the basis of F_{ST} values. The first axis explains 31.8% of the variation, and the second, 23.56% of the variation.

except GAV and MEN, deviated from Hardy-Weinberg equilibrium (HWE). Effective population size estimates were relatively low, especially for MEN (N_e = 4.9), LIO (N_e = 0.4) and KRY (N_e = 4.9). In contrast, BON, TOG, KOR and NEA showed a theoretical infinite value.

Population Differentiation in *C. caespitosa*

The global value of F_{ST} revealed a low but significant level of genetic differentiation (F_{ST} global = 0.043, p = 0). Pairwise F_{ST} values ranged from zero for CDP vs. GAV, CDP vs. COL and CDP vs. PUN (that is, among the western populations) to 0.134 for COL vs. LIO (i.e., between one western and one eastern population) (**Figure 2**; **Table 4**). The standardised F_{ST} value for COL vs. LIO was 0.279. Overall, three ranges of F_{ST} values were detected: from 0 to 0.02 (usually below 0.01) for within region comparisons, 0.02 to 0.06 for comparisons between populations from different regions and 0.06 to 0.13 for comparisons between any of the populations and the Cypriot ones (LIO and KRY), including between them. Notably, we observed relatively low F_{ST} values between the Greek population (NEA) and the Ionian (POC, F_{ST} = 0.039) or the Sardinian (PAL, F_{ST} = 0.037) ones, and high values between KOR and the other Adriatic populations.

The first two axes of the PCoA explained 55.36% of the variation in F_{ST} (**Figure 3**). In this analysis, the populations grouped according to the three main divisions of the Mediterranean basin (western, central and eastern). The western populations from the Spanish peninsular coasts (CDP, PUN and COL) grouped with those from the Balearic Formentera Islands (ESP and GAV). Populations located on the eastern side of the western sub-basin (PAL, BON and MEN) were also differentiated mainly by the first axis. The Adriatic populations, representing a distinctive sub-basin in the middle of the Mediterranean Sea, formed the second group, though with a clear separation between the northern populations of KOR and POR and the rest of the Adriatic populations. With respect to the eastern populations, LIO and KRY in Cyprus were separated from NEA in Greece by the second axis, though all three were clearly separated from one another.

To test whether an IBD pattern was present, we performed a Mantel test with the entire data set and with each of the three groups detected by the STRUCTURE analysis (see below). A pattern of IBD was only observed when all of the populations were analysed together (R^2 = 0.28; p = 0.001). No significant association between genetic differentiation (F_{ST}) and geographic distance was observed when the populations were analysed separately according to sub-basin (**Supplementary Figure S1**).

In the STRUCTURE analyses, three genetically differentiated clusters were detected with Evanno's K method (**Figure 4A**). The

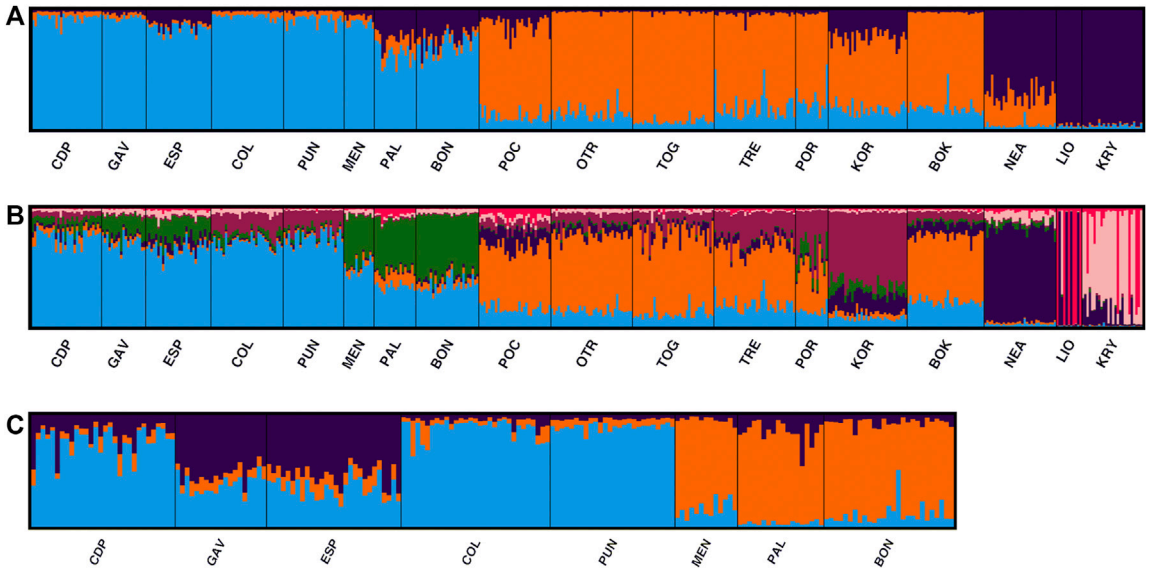


FIGURE 4 | STRUCTURE results selected by Clumpak for the 18 analysed populations of *C. caespitosa* for **(A)** $K = 3$ and **(B)** $K = 7$. **(C)** Results for only the western Mediterranean populations for $K = 3$.

results based on Pritchard's method, which measures the probability (ln (Pr (X/*K*)), showed $K = 9$ as the best value. Under Puechmaille's method, the populations of *C. caespitosa* were divided into seven clusters (**Figure 4B**). Results for $K = 3$ showed a clear division between the three Mediterranean sub-basins (western, central and eastern). The Tyrrhenian (PAL) and Ligurian (BON) populations showed a certain degree of admixture with Espardell Island (ESP) in Formentera. Other populations also showed admixture including KOR in the Adriatic Sea and NEA in the eastern Mediterranean. The results obtained with Puechmaille's method ($K = 7$), though generally similar to those with Evanno's method, showed a substructure within the basins. For instance, in the western Mediterranean, MEN, PAL and BON showed a high percentage of allocation to a particular cluster (besides the most frequent one in the western sub-basin). In addition, the singularity of KOR within the Adriatic Sea was evident, and LIO, NEA and KRY could each be distinguished as a separate group. When we considered $K = 9$, the results remain almost the same as those with a $K = 7$. Analysis of only the western populations revealed three sub-clusters: one differentiating MEN, PAL and BON; another differentiating COL and PUN and a third consisting of GAV and ESP, which are characterised by different degrees of admixture (**Figure 4C**). Cabo de Palos (CDP) showed similarities to both the COL and PUN and GAV and ESP sub-clusters. Supporting the general results of these analyses, the DAPC showed that the Adriatic and the western Mediterranean populations formed two separate groups, and that the Cypriot populations (LIO and KRY) and NEA were separate from each other and from the other groups (**Figure 5A**). Moreover, the DAPC of only the western individuals showed some distinction between the island localities MEN and PAL and the Ligurian one (BON) (**Figure 5B**).

The AMOVA of the three main groups detected by STRUCTURE (western, central and eastern Mediterranean) revealed that 94.67% of the genetic variation originated within populations, with only 2.84% of the variance being attributed to differences among groups. The lowest percentage of variation observed (2.49%) was among populations within groups. Similar results were obtained when seven groups were considered, with 94.75% of the genetic variation originating within populations (**Table 5**).

In the migration analysis, high rates of first generation migrants were detected (**Table 6**). The populations with the highest numbers of migrants were CDP (10 migrants), PUN (6), POC (9) and BOK (6). The most important sink populations were ESP (8) and KOR (6). Migrant flow was most important at an intraregional level, with new individuals coming from populations within the same basin. Given that migrant detection analysis using data comprised of low F_{ST} values may be imprecise, we used the STRUCTURE results for the assignment analyses (**Supplementary Table S1**). These results showed that the individuals from the Formentera Islands were, in part, assigned to one of the same genetic cluster as those from Menorca Island and the Italian populations (PAL and BON). The population from the Ionian Sea (POC) showed genetic admixture with all three Mediterranean sub-basins. In the Adriatic Sea, the northern populations TRE and POR were assigned to one of the same genetic clusters as KOR, the most differentiated location. Between the Cypriot populations, LIO showed putative migrants from KRY.

The results of the BARRIER analysis revealed six supported barriers to gene flow (bootstrap value >95) for *C. caespitosa* in the Mediterranean basin. In the western Mediterranean, two main barriers appear to be present: one separating the north-eastern Spanish populations (COL and PUN) from those in the

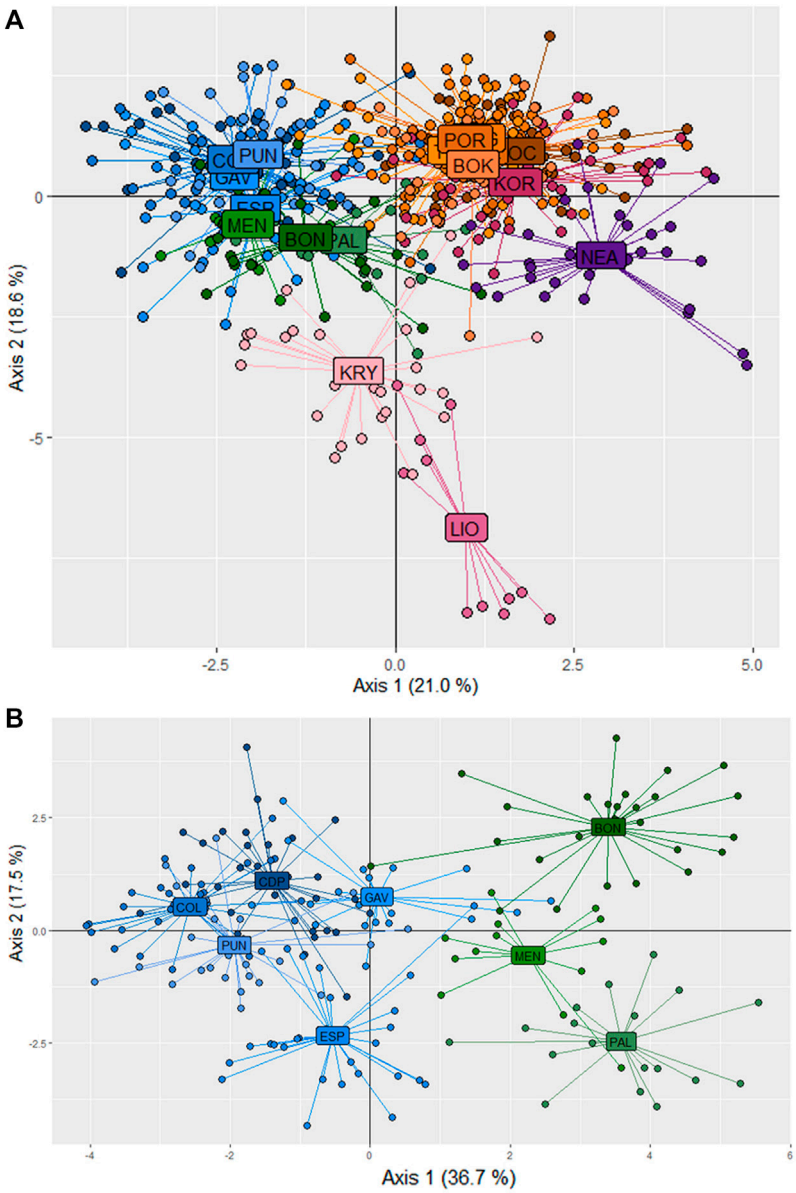


FIGURE 5 | Population differentiation in *C. caespitosa* according to a DAPC analysis of **(A)** all populations or **(B)** only the western Mediterranean populations.

TABLE 5 | AMOVA for the 18 analysed populations of *C. caespitosa* considering (A) three or (B) seven groups. *p* values for all of the results in both analyses were significant ($p < 0.0001$).

Source of Variation	d.f	Sum of Squares	Variance Components	Percentage of Variation
(A) Among groups	2	96.224	0.12891	2.84
Among populations within groups	15	153.617	0.11335	2.49
Within populations	936	4026.192	4.30149	94.67
Total	953	4276.034	4.54376	—
(B) Among groups	6	179.889	0.19975	4.40
Among populations within groups	11	69.953	0.03848	0.85
Within populations	936	4026.192	4.30149	94.75
Total	953	4276.034	4.53972	—

TABLE 6 | First generation migrant test for *C. caespitosa*. For each population (see acronyms in **Table 1**), individuals are presented according to their sampling site in rows, and in columns, to their origin.

	Origin																	
POP	**CDP**	**GAV**	**ESP**	**COL**	**PUN**	**MEN**	**PAL**	**BON**	**POC**	**OTR**	**TOG**	**TRE**	**POR**	**KOR**	**BOK**	**NEA**	**LIO**	**KRY**	**Total**
CDP	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
GAV	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	2
ESP	2	2	—	—	1	1	—	—	—	—	—	—	—	—	1	1	—	—	8
COL	2	—	—	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	5
PUN	—	—	—	1	—	—	—	1	1			1	—	—	—	—	—	—	4
MEN	2	1	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	5
PAL	—	—	—	—	1	—	—	1	—	—	—	—	—	1	—	—	—	—	3
BON	2	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	4
POC	—	—	—	—	—	—	—	—	—	—	1	1	—	1	1	—	—	—	4
OTR	—	—	—	—	—	—	—	—	2	—	—	—	—	1	1	—	—	—	4
TOG	—	—	—	—	—	1	—	—	—	1	—	1	—	—	—	1	—	—	4
TRE	—	—	—	1	1		—	—	1	—	—	—	1	—	1	—	—	—	5
POR	—	—	—	—	—	—	—	—	1	—	2	1		—	—	—	—	—	4
KOR	—	—	—	—	1	—	—	—	1	1		1	2	—	—	—	—	—	6
BOK	—	—	—	—	—	—	—	—	—	1	1		1	—	—	—	—	—	3
NEA	1	—	—	—	—	—	—	1	1	—	—	—	—	—	1	—	—	—	4
LIO	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
KRY	—	—	—	—	—	—	—	—	2	—	—	—	—	—	1	—	1	—	4
Total	10	4	—	2	6	3	1	5	9	4	5	5	4	3	6	2	1	1	

TABLE 7 | Bottleneck analysis results for the analysed populations of *C. caespitosa*.

	Sign Test			Wilcoxon Test			Mode Shift
	IAM	**SMM**	**TPM**	**IAM**	**SMM**	**TPM**	
CDP	0.23112	**0.00847**	0.53546	**0.02693**	0.98288	0.46614	Normal
GAV	**0.00251**	0.31129	0.05764	**0.01184**	0.63314	**0.04488**	Normal
ESP	**0.04321**	0.31652	0.47442	**0.03327**	0.90181	0.28992	Normal
COL	0.22302	**0.02989**	0.33052	0.23415	0.99088	0.73869	Normal
PUN	0.10110	**0.02814**	0.33920	**0.02693**	0.98075	0.43252	Normal
MEN	0.31287	0.40186	0.35531	0.06618	0.79812	0.51839	Normal
PAL	0.13159	0.53566	0.30376	**0.01184**	0.50000	0.07076	Normal
BON	0.05388	0.29483	0.51086	**0.01712**	0.86774	0.15190	Normal
POC	**0.03860**	0.07945	0.52037	**0.00912**	0.81539	0.14186	Normal
OTR	0.11330	**0.00793**	0.51491	0.14186	0.99088	0.61699	Normal
TOG	0.31066	**0.01393**	0.12158	**0.04919**	0.98474	0.64405	Normal
TRE	0.10957	**0.03099**	0.51748	**0.02414**	0.92298	0.41586	Normal
POR	0.46105	0.06859	0.18705	0.09819	0.95512	0.56748	Normal
KOR	0.12798	**0.00784**	0.51569	**0.04937**	0.97586	0.63314	Normal
BOK	0.28149	0.26926	0.51297	0.14186	0.89393	0.69527	Normal
NEA	0.14254	**0.00677**	0.30314	0.06487	0.99720	0.64906	Normal
LIO	**0.00052**	0.13288	0.09412	**0.00019**	**0.01248**	**0.00314**	**Shifted**
KRY	0.42720	0.13931	0.51549	**0.03327**	0.94581	0.50000	Normal

The two statistical tests, Sign and Wilcoxon sign-rank, were conducted under three mutation models: infinite alleles (IAM), stepwise mutation (SMM) and two-phase mutation (TPM). Significant values are in bold. Mode shift is also indicated.

Formentera Islands (GAV and ESP), and another separating ESP (Formentera) from the Menorca population (MEN) (corresponding to the Balearic Front and Mallorca Channel respectively, see **Figure 1**). A barrier also appears to separate the northernmost western populations (BON and PUN) from one another. A barrier that runs between Sicily and North Africa (Sicilian Channel) separates all of the western populations from the central and eastern ones (**Figure 1**). In the Adriatic Sea, a north-south barrier appears to divide the eastern coastal populations from the western ones, except in the north (for POR and TRE). Between the eastern and the central populations, a supported barrier was only observed between OTR and NEA (Aegean Front in **Figure 1**). Finally, in relation to population bottlenecks, only LIO displayed a mode shift, indicating it has experienced a recent bottleneck (**Table 7**).

DISCUSSION

Genetic Variability

Cladocora caespitosa displays both sexual and asexual reproductive modes, and local populations display differences

in the prevalence of these modes. Eight populations across the Mediterranean Sea (three western, four Adriatic and one eastern) appear to undergo sexual reproduction exclusively, as no evidence of clones was found. Sexual reproduction is also dominant in most of the other populations as only a low level of asexual reproduction was observed in them. These results are consistent with those of a previous study showing that sexual reproduction is predominant in *C. caespitosa* (Kružić et al., 2008). By contrast, asexual reproduction is dominant in CRE and LIO (>60% clones), and to a lesser extent, in MEN (38.1% clones). These values are comparable to the level of clonality shown by coral species that reproduce mostly by fragmentation, such as *Pavona clavus* or *Acropora valida* (Aranceta-Garza et al., 2012). The high rate of asexual reproduction in the LIO population was recently related to local natural or anthropogenic disturbances (López-Márquez et al., 2021). This phenomenon has also been observed in other scleractinian corals, such as in populations of *Pocillopora verrucosa* after tropical storms (Aranceta-Garza et al., 2012). Currently, we do not know the specific circumstances that have led to the high level of asexual reproduction in the CRE and MEN populations, though we hypothesise that they also experienced local disturbances that resulted in high mortality rates. In Crete, for instance, tourism has greatly impacted the coastlines (Tsilimigkas et al., 2020), which may contribute to the high level of asexual reproduction observed in *C. caespitosa* in this area. In Menorca, colonies are located in a sheltered place with a low swell (J. Templado, personal observation), thus bioerosion may be weakening these colonies, leading to polyp detachment. The small estimated effective size of MEN also suggests this locality may have been affected by environmental disturbances that could have triggered asexual reproduction in these colonies, as has been observed in other coral species (Lirman, 2000). Therefore, it is accepted the hypothesis that the clonal structure varies across its geographic range due to different disturbances affecting the sexual/asexual reproductive ratio.

Despite having a low level of genotypic diversity (G_o/G_e), the genetic diversity (H_o) of MEN was similar to nearly all of the other analysed populations of *C. caespitosa*, and analogous to those reported for other tropical corals (Baums, 2008). This finding suggests that the level of sexual reproduction in this population is enough to maintain its diversity (Bengtsson, 2003). In concordance with other studies (Baums, 2008; Polato et al., 2010; Casado-Amezúa et al., 2012), we observed positive values for the inbreeding index, a heterozygosity deficit in most populations and deviation from HWE in all populations, except GAV and MEN. Deviation from HWE in marine invertebrate populations may be associated with their sexual reproductive mode. For instance, inbreeding is widespread in most marine benthic invertebrates with limited dispersal capabilities (Riesgo et al., 2016). In *C. caespitosa*, fertilisation takes place in the surrounding water of the broadcasting colonies without sexual selection; therefore, inbreeding is prevalent in this sessile species (Addison and Hart, 2005).

Population Structure

Given the life history traits of *C. caespitosa*, such as its supposedly low dispersal capability and high self-recruitment rates (Kružić et al., 2008; López-Márquez et al., 2019; this study in which a rate greater than 80% was estimated), we hypothesised that the species would have a strong population structure with a high level of differentiation. In contrast, and rejecting this hypothesis, we found that the populations studied had, in general, relatively low F_{ST} values among some nearby populations, indicating a low level of differentiation. However, we did observe a regional structure in the western Mediterranean according to the F_{ST} values: populations from the peninsular coasts of Spain (CDP, PUN and COL) and the Balearic island of Formentera (GAV and ESP) formed a single group, whereas the Menorca island population (MEN) was genetically more similar to the populations in the Tyrrhenian (PAL) and Ligurian (BON) sub-basins (see **Figure 2**). This substructure is likely mediated by dispersal along the coastline and some resistance to larval movement in the open sea since lower F_{ST} values were observed between the coastal populations than between these and the island ones. Similar results were found in the Adriatic and adjacent Ionian populations (López-Márquez et al., 2019) and in the eastern Mediterranean populations (López-Márquez et al., 2021), where the gene flow resulted to be mainly driven by the shoreline configuration and sea surface current which likely enhanced the dispersal of planulae among populations. Genetic differentiation of the Croatian locality KOR from its neighbours in the Adriatic Sea was evidenced by the relatively high F_{ST} values observed among these populations. López-Márquez et al. (2019) showed the same results with a different set of markers and inferred that this was due to the influence of a north-south barrier. Although we also observed this latitudinal separation with our data set, the inclusion of POR, the northernmost Adriatic population, suggests this differentiation is not as strong as previously observed. The F_{ST} values shown between POR and the other Adriatic populations were not as high as those between KOR and these populations. The north-south differentiation observed in the Adriatic Sea is likely mediated by dispersion along the coastline and some dispersal events through subgyres that promote connectivity between the southern populations. Aside from its northern location, KOR stands out as a differentiated cluster. With respect to the eastern Mediterranean populations (KRY and LIO, in the Levantine Sea, and NEA, in the Aegean Sea), pairwise F_{ST} comparisons indicate that all are differentiated from one another; however, in the STRUCTURE analysis (for $K = 3$), they form a single cluster.

The population structure resulting from the various analyses (i.e., STRUCTURE, DAPC and PCoA) coincides with the three main divisions of the Mediterranean Sea (western, central and eastern Mediterranean). According to the Mantel test results, this is likely due to geographic distance (IBD) along the eastern–western distribution axis of the Mediterranean populations (see **Supplementary Figure S1**).

In addition to geographic distance, barriers to gene flow have likely influenced the current population structure of *C. caespitosa*. The barriers we identified largely coincide with those found in other population genetic studies (Pascual et al., 2017; Constantini et al., 2018). For this reason, the results suggesting the presence of long-distance first generation migrants in NEA and KRY were unexpected. One possible explanation could be the evolutionary

convergence of alleles from different locations, which would lead to low pairwise differentiation and, consequently, indicate the presence of migrants. The fact that the standardised F_{ST} values between the eastern Mediterranean and Adriatic and Ionian populations are higher than for other pairwise comparisons supports this premise. Otherwise, anthropogenic dispersal should be taken into account (Radziejewska et al., 2006): human-mediated transport, for instance, through lithic anchors or ballast stones in Phoenician times could have also led to this surprising result.

Regarding the intra-regional substructure detected by the STRUCTURE and assignment analyses, gene flow is not mediated by distance, but rather by ocean currents and dispersal along the coastlines. Consistent with this, the Balearic Front (BF) and the Mallorca Channel (MC, Ruiz et al., 2009) appear to act as barriers dividing CDP and the two Formentera populations (GAV and ESP) from PUN and COL. However, evidence of migrants and a low to moderate level of genetic differentiation indicate that these barriers are permeable. The NBF is seasonal, and it depends on the strength of the Northern Current (Font et al., 1995). As such, during autumn and winter, its intensity decreases and the barrier is less strong. This event almost matches the spawning period of *C. caespitosa* in this region (Kersting et al., 2013b), which would allow for some level of gene flow between populations. Another ocean front of importance is the Northern Tyrrhenian Gyre (Poulain et al., 2012), which facilitates connectivity between the populations from the Tyrrhenian and Ligurian seas (PAL and BON, respectively). On the other hand, connectivity among the eastern populations seem to be influenced mainly by a pattern of isolation by environment, consistent with the results shown by López-Márquez et al. (2021).

Lastly, although we only obtained five unique multilocus genotypes for CRE, leading us to exclude it from further analysis, our results on this population's genetic structure (**Supplementary Figure S2**) were potentially interesting. Greater sampling effort of this area, however, is needed to confirm this hypothesis.

As discussed here, various factors affect the genetic differentiation of the different populations of *C. caespitosa* in the Mediterranean. The population structure of this coral mainly appears to be correlated with the division of the main Mediterranean sub-basins, and the barriers between them, together with favoured dispersal along coastlines. Despite the low dispersal ability, high rate of self-recruitment and estimated low effective population sizes of *C. caespitosa*, the occasional dispersive events that seem to occur between regions are sufficient to maintain the species' considerable genetic diversity. Under suitable conditions and habitat continuity without barriers to gene flow, connectivity between distant colonies of *C. caespitosa* could be high. In fact, the two localities on the peninsular coast of Spain (CDP and PUN), which are separated by 400 km, did not present any genetic differentiation (pairwise F_{ST} value = 0).

Conservation Implications

Despite its importance in conservation policies, knowledge of the genetic diversity and connectivity patterns of species is often not considered by policymakers. This type of information is especially important for the conservation of endangered structural species. Our results on the genetic structure and population connectivity of *C. caespitosa*, whose populations are in alarming decline, highlight key aspects that should be included in the conservation decision making process for this species. These aspects include the relatively low dispersal ability and the high rate of self-recruitment of the species, the importance of sporadic dispersal events to maintain diversity and the level of clonality of some populations.

The regression of populations of *C. caespitosa* is still in progress (Rodolfo-Metalpa et al., 2005), and the high frequency of mortality events in this coral over the last decades possibly exceeds its recovery potential because of its low recruitment rates (Kersting et al., 2014a). Therefore, understanding the general pattern of connectivity among populations of this coral, particularly in light of its discontinuous distribution (as isolated colonies or, more rarely, as beds or banks) and low dispersal ability, becomes increasingly important. Known beds and banks of the coral should be considered as source populations from which larvae can be occasionally exported to other areas. Conversely, populations comprised of a few scattered colonies for which successful fertilisation may prove highly difficult can serve as potential recipient populations. For this reason, strict protection of the marine areas where beds and banks of *C. caespitosa* are known to be present should be a top priority. According to the available information from the literature and expert observations describing the abundance and morphology of colonies (compiled by Chafaoui et al., 2017), only 31 localities across the entire Mediterranean basin have reported bioherms of the species. Although in this study we cannot appraise whether the current level of protection in established marine protected areas is sufficient for the conservation of the species, we strongly recommend an increase in the coverage of this emblematic species within these areas and measures to preserve it from major threats, particularly those that are mediated by human activities. For instance, although extreme climatic events, such heat waves, appear to be a major threat to this coral (Kersting et al., 2013a; Jiménez et al., 2016), other factors posing a threat include the spread of mucilage and invasive algae (Kersting et al., 2014b), eutrophication due to sewage discharge or fish farming (Kružić and Požar-Domac, 2007), trawling, anchorage and high sedimentation rates caused by dredging or extreme storms (Casado de Amezua et al., 2015; Chefaoui et al., 2017; López-Marquez et al., 2021). In addition to the conservation of the known coral beds, prospecting studies of other Mediterranean localities should be conducted to identify as yet unknown populations of the coral.

Besides local impacts, global changes (sea warming and acidification) also threaten *C. caespitosa*. Global threats such heat waves or extreme storms interact with local ones such as pollution or coastal degradation, leading to cumulative impacts that are particularly high in coastal ecosystems (Micheli et al., 2013). Although local mitigation strategies cannot directly protect populations from the impact of global threats, they are often the only feasible way to reduce the synergy between the

different types of threats, and to help preserve marine coastal ecosystems (Templado, 2014). In this context, protection against local impacts is of the utmost importance.

Another aspect to consider in the conservation of *C. caespitosa* is the role of asexual reproduction, and its effect on the genetic diversity of populations. For sessile organisms that undergo external fertilisation, asexual reproduction may be the local means of proliferation for populations with a small population size and/or low gamete density ("Allee effect"; Courchamp et al., 2008). Clonal growth, therefore, may allow the species to persist through periods of low sexual recruitment (Lasker and Coffroth, 1999); however, it may also lead to reduced genotypic diversity and, as a result, higher susceptibility to environmental changes (Reusch et al., 2005). From a conservation point of view, knowledge of the genetic and genotypic diversity patterns of structural species with a high potential for asexual reproduction is critical: populations with a high level of genotypic diversity may be better able to withstand environmental changes or extreme climatic events (Reusch et al., 2005) and, conversely, those with a low level of genetic diversity may be more vulnerable to pathogens and parasites (King and Lively, 2012). For *C. caespitosa*, we found that the ratio of clonal to sexual recruitment is highly variable between localities, and is likely a consequence of a complex interplay of various impacts. Future studies should focus on determining the factors that affect this ratio and how different impacts interact synergistically on it, as well as on the historical causes for the decline of bioherms of the species. Increased clonality may negatively impact the adaptive potential of the species under the current disturbance regime; therefore, it should be closely monitored.

In summary, enforcement of conservation measures that aim to protect the genetic diversity of endangered structural species such *C. caespitosa* is essential. The results provided here, combined with those from a recent study on another endangered Mediterranean coral, *Astroides calycularis* (Ledoux et al., 2021), may help build a highly relevant framework in which to study the evolution of Mediterranean marine diversity as it faces the increasing warming of the basin waters.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Raw Data and **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MR and VL-M performed the analyses and wrote the manuscript, contributing equally to this study. VL-M, JT and AM conceived the research. VL-M, JT and MT conducted field work. AM was responsible for resource acquisition. All authors reviewed and approved the manuscript.

FUNDING

This research was funded by the Spanish Ministry of Economy and Competitiveness (Grant reference: CTM 2014-57949-R), the Spanish Ministry of Science and Innovation (Grant reference: PID 2019-108644 GB-I00), the project Coralien funded by the Fundación Biodiversidad (Spanish Ministry of Ecological Transition and Demographic Challenge), and the European project CoCoNET "Towards COast to COast NETworks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential" from the VII FP of the European Commission (grant agreement n° 287844). ENALIA's research program and *Nephila* Works Ltd. also provided financial support.

ACKNOWLEDGMENTS

Coral samples were collcected by Iván Acevedo, Paula Rodríguez-Flores, Mar Soler, Lorenzo Zane, Louis Hadjioannou, Carlos Jimenez, Marta Calvo, Diego Kurt, Ángel Luque, Paolo Montagna, Eduard Serrano, Anna M. Addamo, Pilar Casado de Amezúa, Francesca Benzoni, Fabrizio Stefani and Antheus s. r.l. We thank the Fisheries Research Institute (Nea Peramos, Greece) for facilitating sample collection and the acquisition of CITES permits. Thanks to the Computational Biogeography and photography laboratories of the MNCN and to Melinda Modrell for the revision of the language. This is ISMAR-CNR Bologna scientific contribution n. 2068.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.889672/full#supplementary-material

REFERENCES

Addison, J. A., and Hart, M. W. (2005). Spawning, Copulation and Inbreeding Coefficients in Marine Invertebrates. *Biol. Lett.* 1, 450–453. doi:10.1098/rsbl.2005.0353

Aguirre, J., and Jiménez, A. P. (1998). Fossil Analogues of Present-Day *Cladocora caespitosa* Coral Banks: Sedimentary Setting, Dwelling Community, and Taphonomy (Late Pliocene, W Mediterranean). *Coral Reefs* 17, 203–213. doi:10.1007/s003380050119

Amorosi, A., Antonioli, F., Bertini, A., Marabini, S., Mastronuzzi, G., Montagna, P., et al. (2014). The Middle-Upper Pleistocene Fronte Section (Taranto, Italy): an Exceptionally Preserved Marine Record of the Last Interglacial. *Glob. Planet. Change* 119, 23–38. doi:10.1016/j.gloplacha.2014.04.007

Aranceta-Garza, F., Balart, E. F., Reyes-Bonilla, H., and Cruz-Hernández, P. (2012). Effect of Tropical Storms on Sexual and Asexual Reproduction in Coral *Pocillopora verrucosa* Subpopulations in the Gulf of California. *Coral Reefs* 31, 1157–1167. doi:10.1007/s00338-012-0941-9

Baums, I. B. (2008). A Restoration Genetics Guide for Coral Reef Conservation. *Mol. Ecol.* 17, 2796–2811. doi:10.1111/j.1365-294X.2008.03787.x

Baums, I. B., Miller, M. W., and Hellberg, M. E. (2006). Geographic Variation in Clonal Structure in A Reef-Building Caribbean Coral, *Acropora palmata*. *Ecol. Monogr.* 76, 503–519. doi:10.1890/0012-9615(2006)076[0503:GVICSI]2.0.CO;2

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. (2004). GENETIX4. 05, logiciel sous Windows TM pour la génétique des populations. *Lab. génome, Popul. Interact. CNRS Umr* 5000, 1996–2004.

Bellan-Santini, D., Bellan, G., Bitar, G., Harmelin, J.-G., and Pergent, G. (2007). *Handbook for Interpreting Types of Marine Habitat for the Selection of Sites to Be Included in the National Inventories of Natural Sites of Conservation Interest*. Tunis: UNEP/MAP-RAC/SPA.

Bengtsson, B. O. (2003). Genetic Variation in Organisms with Sexual and Asexual Reproduction. *J. Evol. Biol.* 16, 189–199. doi:10.1046/j.1420-9101.2003.00523.x

Bianchi, C. N., and Morri, C. (2000). Marine Biodiversity of the Mediterranean Sea: Situation, Problems and Prospects for Future Research. *Mar. Pollut. Bull.* 40, 367–376. doi:10.1016/S0025-326X(00)00027-8

Bosio, S., Di Cencio, A., Coletti, G., Casati, S., and Collareta, A. (2021). Exceptionally Preserved Coral Bank and Seagrass Meadow from the Lower Pleistocene of Fauglia (Tuscany, Italy). *Alp. Medit. Quat.* 34, 1–17. doi:10.26382/AMQ.2021.15

Botsford, L. W., Hastings, A., and Gaines, S. D. (2001). Dependence of Sustainability on the Configuration of Marine Reserves and Larval Dispersal Distance. *Ecol. Lett.* 4, 144–150. doi:10.1046/j.1461-0248.2001.00208.x

Casado-Amezúa, P., García-Jiménez, R., Kersting, D. K., Templado, J., Coffroth, M. A., Merino, P., et al. (2011). Development of Microsatellite Markers as a Molecular Tool for Conservation Studies of the Mediterranean Reef Builder Coral *Cladocora caespitosa* (Anthozoa, Scleractinia). *J. Hered.* 102, 622–626. doi:10.1093/jhered/esr070

Casado-Amezúa, P., Goffredo, S., Templado, J., and Machordom, A. (2012). Genetic Assessment of Population Structure and Connectivity in the Threatened Mediterranean Coral *Astroides calycularis* (Scleractinia, Dendrophylliidae) at Different Spatial Scales. *Mol. Ecol.* 21, 3671–3685. doi:10.1111/j.1365-294X.2012.05655.x

Casado-Amezúa, P., Kersting, D. K., Templado, J., and Machordom, A. (2014). Regional Genetic Differentiation Among Populations of *Cladocora caespitosa* in the Western Mediterranean. *Coral Reefs* 33, 1031–1040. doi:10.1007/s00338-014-1195-5

Casado de Amezua, P., Kersting, D., Linares, C. L., Bo, M., Caroselli, E., Garrabou, J., et al. (2015). *Cladocora caespitosa. The IUCN Red List of Threat. Species*, e.T133142A75872554. doi:10.2305/IUCN.UK.2015-2.RLTS.T133142A75872554.en

Chapuis, M.-P., and Estoup, A. (2007). Microsatellite Null Alleles and Estimation of Population Differentiation. *Mol. Biol. Evol.* 24, 621–631. doi:10.1093/molbev/msl191

Chefaoui, R. M., Casado-Amezúa, P., and Templado, J. (2017). Environmental Drivers of Distribution and Reef Development of the Mediterranean Coral *Cladocora caespitosa*. *Coral Reefs* 36, 1195–1209. doi:10.1007/s00338-017-1611-8

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., et al. (2010). The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLoS ONE* 5, e11842. doi:10.1371/journal.pone.0011842

Cornuet, J. M., and Luikart, G. (1996). Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks from Allele Frequency Data. *Genetics* 144, 2001–2014. doi:10.1093/genetics/144.4.2001

Costantini, F., Ferrario, F., and Abbiati, M. (2018). Chasing Genetic Structure in Coralligenous Reef Invertebrates: Patterns, Criticalities and Conservation Issues. *Sci. Rep.* 8, 5844. doi:10.1038/s41598-018-24247-9

Courchamp, F., Berec, L., and Gascoigne, J. (2008). *Allee Effects in Ecology and Conservation*. Oxford: Oxford University Press.

Cowen, R. K., and Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. *Annu. Rev. Mar. Sci.* 1, 443–466. doi:10.1146/annurev.marine.010908.163757

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M., and Freimer, N. B. (1994). Mutational Processes of Simple-Sequence Repeat Loci in Human Populations. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3166–3170. doi:10.1073/pnas.91.8.3166

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimatorv2: Re-implementation of Software for the Estimation of Contemporary Effective Population Size (Ne) from Genetic Data. *Mol. Ecol. Resour.* 14, 209–214. doi:10.1111/1755-0998.12157

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (Version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evol. Bioinform Online* 1, 117693430500100. doi:10.1177/117693430500100003

Foll, M. (2012). BayeScan v2.1 User Manual. Available at: http://www.cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf.

Font, J., Garcia-Ladona, E., and Gorriz, E. G. (1995). The Seasonality of Mesoscale Motion in the Northern Current of the Western Mediterranean: Several Years of Evidence. *Oceanol. Acta* 18, 207–219.

Gaeta, J., Acevedo, I., López-Márquez, V., Freitas, R., Cruz, R., Maggioni, R., et al. (2020). Genetic Differentiation Among Atlantic Island Populations of the Brown Spiny Lobster *Panulirus echinatus* (Decapoda: Palinuridae). *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 868–881. doi:10.1002/aqc.3297

González-Wangüemert, M., Cánovas, F., Pérez-Ruzafa, A., Marcos, C., and Alexandrino, P. (2010). Connectivity Patterns Inferred from the Genetic Structure of White Seabream (*Diplodus sargus* L.). *J. Exp. Mar. Biol. Ecol.* 383, 23–31. doi:10.1016/j.jembe.2009.10.010

Hadjioannou, L. (2019). Ecological and Biological Responses to Environmental Changes, in the Endemic Scleractinian Coral *Cladocora caespitosa* (Linnaeus, 1767) in Cyprus. Ph. D. Thesis (Nicosia: University of Cyprus).

Hedrick, P. W. (2005). A Standardized Genetic Differentiation Measure. *Evolution* 59, 1633–1638. doi:10.1111/j.0014-3820.2005.tb01814.x

Hellberg, M. E. (2007). Footprints on Water: The Genetic Wake of Dispersal Among Reefs. *Coral Reefs* 26, 463–473. doi:10.1007/s00338-007-0205-2

Holland, L. P., Jenkins, T. L., and Stevens, J. R. (2017). Contrasting Patterns of Population Structure and Gene Flow Facilitate Exploration of Connectivity in Two Widely Distributed Temperate Octocorals. *Heredity* 119, 35–48. doi:10.1038/hdy.2017.14

Jiménez, C., Hadjioannou, L., Petrou, A., Nikolaidis, A., Evriviadou, M., and Lange, M. A. (2016). Mortality of the Scleractinian Coral *Cladocora caespitosa* during a Warming Event in the Levantine Sea (Cyprus). *Reg. Environ. Change* 16, 1963–1973. doi:10.1007/s10113-014-0729-2

Jombart, T. (2008). adegenet: a R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart, T. (2012). *A Tutorial for Discriminant Analysis of Principal Components (DAPC) Using Adegenet 1*, 3-4, 1–43. Available at: https://adegenet.r-forge.r-project.org.

Kersting, D.-K., and Linares, C. (2012). *Cladocora caespitosa* bioconstructions in the Columbretes Islands Marine Reserve (Spain, NW Mediterranean): Distribution, Size Structure and Growth. *Mar. Ecol.* 33, 427–436. doi:10.1111/j.1439-0485.2011.00508.x

Kersting, D. K., Ballesteros, E., De Caralt, S., and Linares, C. (2014b). Invasive Macrophytes in a Marine Reserve (Columbretes Islands, NW Mediterranean): Spread Dynamics and Interactions with the Endemic Scleractinian Coral *Cladocora caespitosa*. *Biol. Invasions* 16, 1599–1610. doi:10.1007/s10530-013-0594-9

Kersting, D. K., Bensoussan, N., and Linares, C. (2013a). Long-term Responses of the Endemic Reef-Builder *Cladocora caespitosa* to Mediterranean Warming. *PLoS ONE* 8, e70820. doi:10.1371/journal.pone.0070820

Kersting, D. K., Casado, C., López-Legentil, S., and Linares, C. (2013b). Unexpected Patterns in the Sexual Reproduction of the Mediterranean Scleractinian Coral *Cladocora caespitosa*. *Mar. Ecol. Prog. Ser.* 486, 165–171. doi:10.3354/meps10356

Kersting, D. K., Teixidó, N., and Linares, C. (2014a). Recruitment and Mortality of the Temperate Coral *Cladocora caespitosa*: Implications for the Recovery of Endangered Populations. *Coral Reefs* 33, 403–407. doi:10.1007/s00338-014-1144-3

King, K. C., and Lively, C. M. (2012). Does Genetic Diversity Limit Disease Spread in Natural Host Populations? *Heredity* 109, 199–203. doi:10.1038/hdy.2012.33

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a Program for Identifying Clustering Modes and Packaging Population Structure Inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi:10.1111/1755-0998.12387

Koukouras, A., Kühlmann, D., Voultsiadou, E., Vafidis, D., Dounas, C., Chintiroglou, C., et al. (1998). The Macrofaunal Assemblage Associated with the Scleractinian Coral *Cladocora caespitosa* (L.) in the Aegean Sea. *Ann. Inst. Océanogr.* 74, 97–114.

Kružić, P., and Požar-Domac, A. (2007). Impact of Tuna Farming on the Banks of the Coral *Cladocora caespitosa* in the Adriatic Sea. *Coral Reefs* 26, 665. doi:10.1007/s00338-007-0237-7

Kružić, P., Žuljević, A., and Nikolić, V. (2008). Spawning of the Colonial Coral *Cladocora caespitosa* (Anthozoa, Scleractinia) in the Southern Adriatic Sea. *Coral Reefs* 27, 337–341. doi:10.1007/s00338-007-0334-7

Lasker, H. R., and Coffroth, M. A. (1999). Responses of Clonal Reef Taxa to Environmental Change. *Am. Zool.* 39, 92–103. doi:10.1093/icb/39.1.92

Li, Y.-L., and Liu, J.-X. (2018). StructureSelector : A Web-Based Software to Select and Visualize the Optimal Number of Clusters Using Multiple Methods. *Mol. Ecol. Resour.* 18, 176–177. doi:10.1111/1755-0998.12719

Linnaeus, C. (1767). "Systema Naturae Per Regna Tria Naturae: Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Diferentiis, Synonymis, Loci," in *Regnum Animale. Holmiae: Laurentii Salvi*. Vol. 1. Part 2. doi:10.5962/bhl.title.68927

Lirman, D. (2000). Fragmentation in the Branching Coral *Acropora palmata* (Lamarck): Growth, Survivorship, and Reproduction of Colonies and Fragments. *J. Exp. Mar. Biol. Ecol.* 251, 41–57. doi:10.1016/S0022-0981(00)00205-7

López-Márquez, V., Cushman, S. A., Templado, J., Wan, H. Y., Bothwell, H. M., Kruschel, C., et al. (2019). Seascape Genetics and Connectivity Modelling for an Endangered Mediterranean Coral in the Northern Ionian and Adriatic Seas. *Landsc. Ecol.* 34, 2649–2668. doi:10.1007/s10980-019-00911-x

López-Márquez, V., Lozano-Martín, C., Hadjioannou, L., Acevedo, I., Templado, J., Jimenez, C., et al. (2021). Asexual Reproduction in Bad Times? The Case of *Cladocora caespitosa* in the Eastern Mediterranean Sea. *Coral Reefs* 40, 663–677. doi:10.1007/s00338-020-02040-3

Luikart, G., Allendorf, F., Cornuet, J., and Sherwin, W. (1998). Distortion of Allele Frequency Distributions Provides a Test for Recent Population Bottlenecks. *J. Hered.* 89, 238–247. doi:10.1093/jhered/89.3.238

Manni, F., Guerard, E., and Heyer, E. (2004). Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Hum. Biol.* 76, 173–190. doi:10.1353/hub.2004.0034

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* 27, 209–220.

Micheli, F., Halpern, B. S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., et al. (2013). Cumulative Human Impacts on Mediterranean and Black Sea Marine Ecosystems: Assessing Current Pressures and Opportunities. *PLoS ONE* 8, e79889. doi:10.1371/journal.pone.0079889

Palumbi, S. R., Sandifer, P. A., Allan, J. D., Beck, M. W., Fautin, D. G., Fogarty, M. J., et al. (2009). Managing for Ocean Biodiversity to Sustain Marine Ecosystem Services. *Front. Ecol. Environ.* 7, 204–211. doi:10.1890/070135

Pascual, M., Rives, B., Schunter, C., and Macpherson, E. (2017). Impact of Life History Traits on Gene Flow: A Multispecies Systematic Review across Oceanographic Barriers in the Mediterranean Sea. *PLoS ONE* 12, e0176419. doi:10.1371/journal.pone.0176419

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* 6, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Peirano, A., Kružić, P., and Mastronuzzi, G. (2009). Growth of Mediterranean Reef of *Cladocora caespitosa* (L.) in the Late Quaternary and Climate Inferences. *Facies* 55, 325–333. doi:10.1007/s10347-008-0177-x

Peirano, A., Morri, C., Bianchi, C. N., Aguirre, J., Antonioli, F., Calzetta, G., et al. (2004). The Mediterranean Coral *Cladocora caespitosa*: A Proxy for Past Climate Fluctuations? *Glob. Planet. Change* 40, 195–200. doi:10.1016/S0921-8181(03)00110-3

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., and Estoup, A. (2004). GENECLASS2: a Software for Genetic Assignment and First-Generation Migrant Detection. *J. Hered.* 95, 536–539. doi:10.1093/jhered/esh074

Pitacco, V., Orlando-Bonaca, M., Mavric, B., and Lipej, L. (2014). Macrofauna Associated with a Bank of *Cladocora caespitosa* (Anthozoa, Scleractinia) in the Gulf of Trieste (Northern Adriatic). *Ann. Ser. Hist. Nat.* 24, 1–14.

Polato, N. R., Concepcion, G. T., Toonen, R. J., and Baums, I. B. (2010). Isolation by Distance across the Hawaiian Archipelago in the Reef-building coral *Porites lobata*. *Mol. Ecol.* 19, 4661–4677. doi:10.1111/j.1365-294X.2010.04836.x

Pons-Fita, A., Verdura, J., Santamaría, J., Kersting, D. K., and Ballesteros, E. (2020). Coexistence of the Reef-Building Coral *Cladocora caespitosa* and the Canopy-Forming Alga *Treptacantha ballesterosii*: Description of a New Mediterranean Habitat. *Sci. Mar.* 84, 263–271. doi:10.3989/scimar.05018.11B

Poulain, P.-M., Menna, M., and Mauri, E. (2012). Surface Geostrophic Circulation of the Mediterranean Sea Derived from Drifter and Satellite Altimeter Data. *J. Phys. Oceanogr.* 42, 973–990. doi:10.1175/JPO-D-11-0159.1

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

Puechmaille, S. J. (2016). The Program Structure does Not Reliably Recover the Correct Population Structure when Sampling Is Uneven: Subsampling and New Estimators Alleviate the Problem. *Mol. Ecol. Resour.* 16, 608–627. doi:10.1111/1755-0998.12512

Radziejewska, T., Gruszka, P., and Rokicka-Praxmajer, J. (2006). A Home Away from Home: A Meiobenthic Assemblage in a Ship's Ballast Water Tank Sediment. *Oceanologia* 48, 259–265.

Rannala, B., and Mountain, J. L. (1997). Detecting Immigration by Using Multilocus Genotypes. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9197–9201. doi:10.1073/pnas.94.17.9197

Raymond, M., and Rousset, F. (1995). GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J. Hered.* 86, 248–249. doi:10.1093/oxfordjournals.jhered.a111573

Reusch, T. B. H., Ehlers, A., Hämmerli, A., and Worm, B. (2005). Ecosystem Recovery after Climatic Extremes Enhanced by Genotypic Diversity. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2826–2831. doi:10.1073/pnas.0500008102

Rice, W. R. (1989). Analyzing Tables of Statistical Tests. *Evolution* 43, 223–225. doi:10.1111/j.1558-5646.1989.tb04220.x

Riesgo, A., Pérez-Portela, R., Pita, L., Blasco, G., Erwin, P. M., and López-Legentil, S. (2016). Population Structure and Connectivity in the Mediterranean Sponge *Ircinia fasciculata* Are Affected by Mass Mortalities and Hybridization. *Heredity* 117, 427–439. doi:10.1038/hdy.2016.41

Rodolfo-Metalpa, R., Peirano, A., Houlbrèque, F., Abbate, M., and Ferrier-Pagès, C. (2008). Effects of Temperature, Light and Heterotrophy on the Growth Rate and Budding of the Temperate Coral *Cladocora caespitosa*. *Coral Reefs* 27, 17–25. doi:10.1007/s00338-007-0283-1

Rodolfo-Metalpa, R., Bianchi, C. N., Peirano, A., and Morri, C. (2005). Tissue Necrosis and Mortality of the Temperate coral *Cladocora caespitosa*. *Ital. J. Zool.* 72, 271–276. doi:10.1080/11250000509356685

Rubio-Portillo, E., Kersting, D. K., Linares, C., Ramos-Esplá, A. A., and Antón, J. (2018). Biogeographic Differences in the Microbiome and Pathobiome of the Coral *Cladocora caespitosa* in the Western Mediterranean Sea. *Front. Microbiol.* 9, 1–11. doi:10.3389/fmicb.2018.00022

Ruiz, S., Pascual, A., Garau, B., Faugère, Y., Alvarez, A., and Tintoré, J. (2009). Mesoscale Dynamics of the Balearic Front, Integrating Glider, Ship and Satellite Data. *J. Mar. Syst.* 78, S3–S16. S3–S16. doi:10.1016/j.jmarsys.2009.01.007

Sala, E., and Knowlton, N. (2006). Global Marine Biodiversity Trends. *Annu. Rev. Environ. Resour.* 31, 93–122. doi:10.1146/annurev.energy.31.020105.100235

Schiller, C. (1993). Ecology of the Symbiotic Coral *Cladocora caespitosa* (L.) (Faviidae, Scleractinia) in the Bay of Piran (Adriatic Sea): I. Distribution and Biometry. *Mar. Ecol.* 14, 205–219. doi:10.1111/j.1439-0485.1993.tb00480.x

Stoddart, J. A., and Taylor, J. F. (1988). Genotypic Diversity: Estimation and Prediction in Samples. *Genetics* 118, 705–711. doi:10.1093/genetics/118.4.705

Templado, J. (2014). "Future Trends of Mediterranean Biodiversity," in *The Mediterranean Sea. Its History and Present Challenges*. Editors S. Goffredo and Z. Dubinsky (New York: Springer), 479–498. doi:10.1007/978-94-007-6704-1_28

Tsilimigkas, G., Rempis, N., and Derdemezi, E. T. (2020). Marine Zoning and Landscape Management on Crete Island, Greece. *J. Coast. Conserv.* 24, 43. doi:10.1007/s11852-020-00757-5

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). MICRO-CHECKER: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Mol. Ecol. Notes* 4, 535–538. doi:10.1111/j.1471-8286.2004.00684.x

Vertino, A., Stolarski, J., Bosellini, F. R., and Taviani, M. (2014). "Mediterranean Corals through Time: From Miocene to Present," in *Mediterranean Corals through Time: From Miocene to Present. The Mediterranean Sea. Its History and Present Challenges*. Editors S. Goffredo and Z. Dubinsky (New York: Springer), 257–274. doi:10.1007/978-94-007-6704-1_14

Villamor, A., Costantini, F., and Abbiati, M. (2014). Genetic Structuring across Marine Biogeographic Boundaries in Rocky Shore Invertebrates. *PLoS ONE* 9, e101135. doi:10.1371/journal.pone.0101135

Waples, R. S., and Do, C. (2008). LDNE: A Program for Estimating Effective Population Size from Data on Linkage Disequilibrium. *Mol. Ecol. Resour.* 8, 753–756. doi:10.1111/j.1755-0998.2007.02061.x

Ward, R. D., Woodwark, M., and Skibinski, D. O. F. (1994). A Comparison of Genetic Diversity Levels in Marine, Freshwater, and Anadromous Fishes. *J. Fish Biol.* 44, 213–232. doi:10.1111/j.1095-8649.1994.tb01200.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Brief Research Report
PUBLISHED 12 September 2022
DOI 10.3389/fgene.2022.955382

OPEN ACCESS

EDITED BY
Shaokui Yi,
Huzhou University, China

REVIEWED BY
Cong Zeng,
Shanghai Jiao Tong University, China
Zhi Zhang,
Minjiang University, China

*CORRESPONDENCE
Qianhong Gu,
gqh@hunnu.edu.cn
qianhong_g@163.com

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 28 May 2022
ACCEPTED 21 July 2022
PUBLISHED 12 September 2022

CITATION
Zhong H, Sun Y, Wu H, Li S, Shen Z, Yang C, Wen M, Chen P and Gu Q (2022), Pleistocene climate and geomorphology drive the evolution and phylogeographic pattern of Triplophysa robusta (Kessler, 1876). *Front. Genet.* 13:955382. doi: 10.3389/fgene.2022.955382

Pleistocene climate and geomorphology drive the evolution and phylogeographic pattern of *Triplophysa robusta* (Kessler, 1876)

Hui Zhong[1], Yaxian Sun[1], Huihui Wu[2], Shengnan Li[1], Zhongyuan Shen[1], Conghui Yang[1], Ming Wen[1], Peng Chen[3] and Qianhong Gu[1,2]*

[1]The State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Sciences, Hunan Normal University, Changsha, China, [2]College of Fisheries, Engineering Technology Research Center of Henan Province for Aquatic Animal Cultivation, Henan Normal University, Xinxiang, China, [3]Xinjiang Uygur Autonomous Region Fishery Research Institute, Urumchi, China

Montane systems provide excellent opportunities to study the rapid radiation influenced by geological and climatic processes. We assessed the role of Pleistocene climatic oscillations and mountain building on the evolution history of *Triplophysa robusta*, a cold-adapted species restricted to high elevations in China. We found seven differentiated sublineages of *T. robusta*, which were established during the Mid Pleistocene 0.87–0.61 Mya. The species distribution modeling (SDM) showed an expansion of *T. robusta* during the Last Glacial Maximum (LGM) and a considerable retraction during the Last Interglacial (LIG). The deep divergence between Clade I distributed in Qinling Mountains and Clade II in Northeastern Qinghai-Tibet Plateau (QTP) was mainly the result of a vicariance event caused by the rapid uplifting of Qinling Mountains during the Early Pleistocene. While the middling to high level of historical gene flow among different sublineages could be attributed to the dispersal events connected to the repetition of the glacial period during the Pleistocene. Our findings suggested that frequent range expansions and regressions due to Pleistocene glaciers likely have been crucial for driving the phylogeographic pattern of *T. robusta*. Finally, we urge a burning question in future conservation projection on the vulnerable cold-adapted species endemic to high elevations, as they would be negatively impacted by the recent rapid climate warming.

KEYWORDS

Qinling mountains, vicariance event, past climate change, phylogeographic pattern, T. robusta

1 Introduction

The process of population isolation and differentiation, the core of micro-evolution, is one of the crucial steps for deciphering speciation, which has long fascinated evolutionary biologists (Rice and Hostert, 1993; Wang et al., 2021). The Earth's complex topographical history often generates opportunities for population isolation and differentiation (Yan et al., 2010; Hou et al., 2014; Wang et al., 2021), especially for montane species. Mountain-building is often associated with tectonic forces, promoting population divergence through vicariance (Hoorn et al., 2010; Smith et al., 2014; Simões et al., 2016; Schramm et al., 2021). However, previous studies suggested that mountain building might not be regarded as the sole driver in the diversification of montane species (Kong et al., 2022). Climate change is also considered a major driver of vicariant speciation and intraspecific divergence and animal radiations (Wang et al., 2012; de Oliveira et al., 2021). Particularly, climatic changes associated with Pleistocene glacial cycles are believed to have caused montane species to shift, expand, or contract along latitudinal or elevational gradients (Wang et al., 2012). Populations adapted to different mountains may have experienced isolations due to the range contractions, or may have experienced connectivity as a result of range expansions following climate fluctuations and geological activity.

The drastic uplift of the Qinghai-Tibet Plateau (QTP) and the resulting Asian monsoons largely reshaped the landscape of local and adjacent regions (Clark et al., 2005; Zhang et al., 2012). For example, the Northeastward growth of the Tibetan Plateau since the late Cenozoic and the Northern Qinling Mountains have been strongly reactivated and uplifted. The Qinling Mountains rapidly uplifted during the Miocene-Quaternary (Dong et al., 2011; Liu et al., 2013; Meng, 2017), and extended more than 1,500 km from West to East in Central China. The diverse topography and varied habitats in Qinling Mountains come into being a biogeographical barrier for wild species (Yan et al., 2010; Huang et al., 2017; Wang et al., 2021). Both the Northeastern QTP and Qinling Mountains are important regions for the evolution of biota (Wang et al., 2012; Feng et al., 2019).

Triplophysa are endemic species in QTP and its adjacent area. It is believed that the origin and evolution of *Triplophysa* are related to the uplift of the QTP (Wang et al., 2016; Li et al., 2017). This group has been one major component of fish fauna on the QTP and the adjacent area and has evolved specific morphological and physiological adaptations to the low oxygen, low pressure, and cold conditions of the plateau (Wu et al., 2020). *T. robusta* is endemic to the Upper Yellow River and the upper reaches of the Jialing River (Feng et al., 2019). However, it was reported that *T. robusta* could be found in the upper reaches of the Yihe River and Luohe River, distributed in the Qinling Mountains, as well as the montane streams in the Southern Taihang Mountains (Wu et al., 2020). Therefore, *T. robusta* represents an ideal model to address the effects of past climatic and geological events on the intraspecific diversification among populations adjacent to and distant from QTP. Hence we used mitochondrial and nuclear genes to estimate the phylogeographic pattern and evolutionary history of *T. robusta*, and to evaluate the conservation status and strategy of this species.

2 Materials and methods

2.1 Samples and sequence data preparation

Ten populations of *T. robusta* in the Eastern Qinling Mountains and four populations in the Southern Taihang Mountains (Figure 1; Table 1) were collected in 2017 and 2018. The genomic DNA was extracted and the nuclear genes rhodopsin (*RH1*) was amplified following Feng et al. (2019). We sequenced 42 novel *RH1* genes with three individuals in each population, and they have been deposited in GenBank (accession number, ON630917-ON630958). Three mitochondrial genes (16S rRNA, *16S*; cytochrome b, *cytb*, cytochrome oxidase I, *COI*, Supplementary Table S1) were obtained from previous studies (Feng et al., 2019; Wu et al., 2020). A total of 237 individuals from 28 localities across the major distribution of *T. robusta* were collected in China (Figure 1; Table 1). Besides, another 122 *RH1* sequences were also retrieved from the previous study (Feng et al., 2019), and a total of 164 *RH1* sequences were used to investigate the populations' relationship among the Northeastern QTP, the Eastern Qinling Mountains, and the Southern Taihang Mountains.

2.2 Phylogenetic analysis

All sequences were aligned by using MAFFT implemented in PhyloSuite v1.2.2. We tested the homogeneity of the combined genes (*COI* + *cytb* + *16S* + *RH1*) in *PAUP and the *p*-value was 0.074 (>0.05). Consequently, the aligned nucleotide sequences *cytb*, *COI*, *16S,* and *RH1* were concatenated to construct the phylogenetic tree by using the maximum likelihood (ML) method implemented in RAxML v8.2.4 with 1,000 bootstrap replicates. Each gene was treated as unlinked since it is not required to have the same individuals analyzed for every genetic marker if partitions are considered as unlinked (Sota and Vogler, 2001; Husemann et al., 2013; Gu et al., 2022). Bayesian inferences (BI) analysis was performed in MrBayes v3.2.6 using the Markov Chain Monte Carlo method with 10 million generations and sampling trees every 1,000 generations. The first 25% of trees were discarded as burn-in with the remaining trees being used for generating a consensus tree. The species of *Triplophysa siluroides* and *Triplophysa strauchii* (Supplementary Table S1) were used as outgroups. The final trees were visualized in FIGTREE v1.4.4.

To further visualize haplotype diversity and distribution of *T. robusta*, we generated a haplotype network using 235 *cytb*

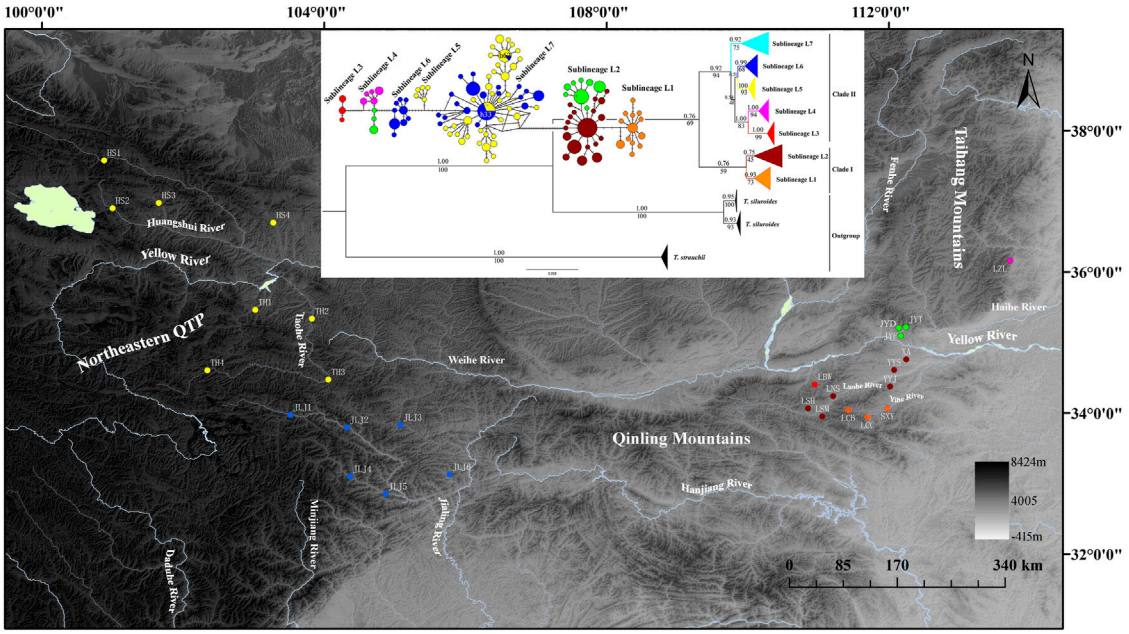


FIGURE 1
Map of sampling locations of *T. robusta*, localities are detailed in Table 1 according to Wu et al. (2020) and Feng et al. (2019), and the phylogenetic tree based on Bayesian inference (values on branches indicate bootstrap proportions from a maximum likelihood analysis and bayesian posterior probabilities), and the median-joining haplotype network based on 235 *cytb* sequences, the size of the circles represents haplotype frequency. Each connecting line represents a single nucleotide substitution, and each little short line represents a mutated position. Different river systems were represented in different colors, consistent with the colors of the haplotype.

sequences and colored each haplotype by the geographic region from which river system it was collected (Table 1). The haplotype networks were constructed using network v. 10 (www.fluxus-engineering.com/sharenet.htm) and applying the median-joining and maximum parsimony options. Seven sublineages were found in the present study, and the diversity indices of each population and sublineage were calculated using DNASP v6.12.03. Two neutrality statistics, Tajima's *D* and Fu's *Fs* were also calculated with 1,000 permutations in ARLEQUIN v. 3.5.2.2. Analysis of molecular variance (AMOVA) was conducted in ARLEQUIN to partition the genetic variance within and among populations. Furthermore, four groups were defined according to the river system and Mountains system (HS1-4 and TH1-4 as Yellow River system in Northeastern QTP; JLJ1-6 as Yangtze River system in Northeastern QTP; LBW, LSH, LSM, LNS, YYJ, YYS, XA, LCB, LCC, and SXY as Yellow River system in Qinling Mountains; the other four populations as River systems in Southern Taihang Mountains) to estimate variance components.

2.3 Demographic history and gene flow

As seven sublineages of *T. robusta* were found in phylogenetic analyses. We used the *cytb* data to generate Bayesian skyline plots (BSP) to study the demographic history of each sublineage alone and the whole combined data set. The appropriate substitution model was determined as GTR with Moldetest v.3.7. We created individual input files for each sublineage and all combined with BEAUti available in the BEAST package. Analyses were run in BEAST v.2.6.7. We used a strict clock with the average evolution rate of 0.68%–0.86% per million years for Cobitidae estimated for *cytb* under the GTR model (Perdices et al., 2014). The program was run for 20 million generations sampling every 2,000 generations. The Bayesian Skyline plots were subsequently generated with TRACER v.1.7.1.

Pairwise migration rates between the seven sublineages were estimated using the *cytb* and a maximum likelihood coalescent approach implemented in MIGRATE-n v 3.6.4; we estimated Θ and *M* (immigration rate/mutation rate) based on F_{ST} values. We ran 10 short chains with a total of 10,000 genealogy samples and three long chains with 1,000,000 samples, following a burn-in of 10,000 samples; four independent runs were performed.

2.4 Divergence time estimation

The divergence time was estimated using a molecular clock approach as implemented in BEAST v2.6.7 with a relaxed clock model. We used a reduced dataset of *cytb*: specimens for which three

10.3389/fgene.2022.955382

TABLE 1 Detailed information for specimens included in this study and diversity indices for each population based on *cytb*.

Pop ID	N	Locations	Latitude	Longitude	Number of genes sequenced				Diversity indices		
					16S	*COI*	*cyt b*	*RH1*	S	Hd	π
HS1	13	Huangshui River	37.578	100.881	13	0	13	13	12	0.936	0.00317
HS2	8	Huangshui River	36.900	101.000	8	0	8	8	18	0.929	0.00501
HS3	3	Huangshui River	36.977	101.654	3	0	3	3	6	1.000	0.00351
HS4	7	Huangshui River	36.697	103.278	7	0	7	7	19	0.810	0.00643
TH1	10	Taohe River	35.460	103.022	10	0	10	10	19	1.000	0.00541
TH2	16	Taohe River	35.335	103.825	16	0	16	16	24	0.975	0.00448
TH3	7	Taohe River	34.473	104.057	7	0	7	7	16	0.929	0.00473
TH4	1	Taohe River	34.604	102.344	1	0	1	1	NA	NA	NA
JLJ1	17	Jialing River	33.973	103.522	17	0	17	17	16	0.926	0.00501
JLJ2	5	Jialing River	33.795	104.325	5	0	5	5	18	1.000	0.00825
JLJ3	3	Jialing River	33.830	105.080	3	0	3	3	12	1.000	0.00702
JLJ4	3	Jialing River	33.099	104.367	3	0	3	3	5	1.000	0.00292
JLJ5	23	Jialing River	32.849	104.872	23	0	23	23	8	0.818	0.00139
JLJ6	6	Jialing River	33.128	105.776	6	0	6	6	14	0.933	0.00427
LZL	9	Lushui River	36.157	113.718	0	9	9	3	4	0.861	0.00120
LBW	6	Dongjian River	34.403	110.952	0	5	5	3	25	0.867	0.00786
LCB	4	Yihe River	34.043	111.423	0	4	4	3	4	0.833	0.00195
LCC	9	Yihe River	33.939	111.697	0	9	9	3	11	0.889	0.00320
SXY	6	Yihe River	34.064	111.980	0	6	6	3	7	0.933	0.00246
LNS	9	Luohe River	34.240	111.211	0	9	9	3	10	0.917	0.00218
LSH	3	Luohe River	34.064	110.852	0	3	3	3	3	0.667	0.00180
LSM	9	Luohe River	33.951	111.055	0	9	9	3	12	0.889	0.00360
XA	16	Luohe River	34.758	112.245	0	16	16	3	4	0.725	0.00116
YYJ	5	Luohe River	34.376	112.016	0	5	5	3	2	0.400	0.00072
YYS	9	Luohe River	34.610	112.073	0	9	9	3	0	0.000	0.00000
JYD	11	Dongyang River	35.203	112.139	0	11	11	3	15	0.764	0.00517
JYF	7	Fengshi River	35.093	112.167	0	7	7	3	13	0.810	0.00523
JYT	12	Tieshan River	35.215	112.240	0	12	10	3	49	0.933	0.02100
Total	237				122	114	234	164	153	0.985	0.01846

N means the number of individuals in each population, S indicates the number of polymorphic sites, Hd indicates haplotype diversity, π indicates nucleotide diversity, NA, means absent.

to six sequences from each population were included. This method resulted in 106 terminals—98 representatives of the 28 populations of *T. robusta* and eight outgroup sequences (*T. siluroides* and *T. strauchii*) (Supplementary Table S2). As fossil evidence for *Triplophysa* was lacking, we used a conservative approach by employing two calibration points from previous studies (Feng et al., 2019; Wu et al., 2020). The most recent common ancestor (TMRCA) of *Triplophysa* dates from the early Miocene, corresponding to the early Miocene QTP uplift (Wu et al., 2020). The prior "speciation: Yule process" tree was used to construct the tree. We ran four independent runs for 20 million generations logging trees every 2000 generations. Convergence was checked with TRACER v. 1.7.1. The maximum credibility tree was created in TreeAnnotator v. 2.6.7 available in the BEAST package.

2.5 Reconstruction of ancestral areas

A biogeographic reconstruction of ancestral areas using BioGeoBEARS was performed in RASP 4.02. Ancestral distributions were reconstructed using 31 *cytb* sequences covering the range of seven sublineages of *T. robusta* (Supplementary Table S3). The 7,500 post-burn-in trees resulting from the BEAST analysis were integrated for inference. Five major areas for *T. robusta* according to our collection localities and the published literature (Wang et al., 2016; Li et al., 2017; Feng et al., 2019; Wu et al., 2020) are depicted as follows: (A) Upper Yellow River system in Northeastern QTP, (B) Jialing River system in Northeastern QTP, (C) Yellow River system in Qinling Mountains, (D) Yellow River system in

Southern Taihang Mountains, and (E) Haihe River system in Southern Taihang Mountains. All six models of geographic range evolution were compared in a likelihood framework. The DIVALIKE + *j* was chosen by comparing Akaike's information criterion and likelihood-ratio tests.

2.6 Species distribution and paleo-distribution 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) *T. robusta* distributions. The occurrence records for *T. robusta* were collected from all published literature about *T. robusta* (Li et al., 2017; Feng et al., 2019; Wu et al., 2020), and a total of 29 unique, geo-referenced and quality-checked occurrence records were finally retained in the present study (Supplementary Table S4).

We extracted current bioclimatic data from the WORLDCLIM dataset (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 (Hasumi and Emori, 2004), and LIG bioclimatic data from Otto-Bliesner et al. (2006) at 30 arc-seconds resolution. The ArcGIS 10.0 (ESRI: Redlands, CA) was used to extract climatic variable layers to include the distribution area (QTP and adjacent area, Qinling Mountains, and Taihang Mountains) of *T. robusta* to improve the predictive power of Maxent models (Anderson and Raza, 2010). All layers were clipped to an extent encompassing the known range of *T. robusta*, as well as adjacent, potentially accessible habitats (30–40° N and 98–116° E). Prior to constructing SDM, the ENMTools v. 1.3 was used to determine which bioclimatic variables were correlated, using R > 0.90 as a cutoff, and fifteen variables (bio1, bio2, bio5, bio6, bio8, bio9, bio10, bio11, bio12, bio13, bio14, bio16, bio17, bio18, and bio19) which were found to be correlated with at least one other variable were removed. The following four predictor variables were retained: bio3, 4, 7, and 15. Using the four Bioclim layers and spatially filtered occurrence data, We ran the Maxent v. 3.4.4 to construct a present-day SDM, and then projected it to the LIG and LGM climate conditions, and used the following settings for the Maxent model: hinge features only, regularization multiplier of 1,10,000 max number of background points, replicate the run type of 10 cross-validations (equivalent to 20% testing), logistic output, 500 maximum iterations, and 0.00001 convergence threshold, and a random seed. Furthermore, we optimized the regularization multiplier by generating models with values from 1 to 10 and chose the value with the best Akaike Information Criterion score. We ran jackknife tests to measure the importance of each bioclimatic variable. Models used 29 records for testing and the four BIOCLIM environmental layers to produce models for present and paleodistributions of *T. robusta*. The logistic output, which was a continuous probability of presence ranging from 0 to 1, was displayed in ARCGIS 10.0.

3 Results

3.1 Phylogeographic pattern of *T. robusta*

Both phylogenetic tree (ML and BI) and haplotype network showed a clear identification of seven sublineages found in *T. robusta* (Figure 1). There was no shared haplotype among the six sublineages (L1-L6) in different river systems and mountains, except for the sublineage L7 in upstream of Yellow River and Jialing River. Therefore, a distinct biogeographic structure with six particular geographical regions was found across the investigation area, two in Northeastern QTP (sublineage L5 and L6), three in Eastern Qinling Mountains (sublineage L1, L2, and L3), and one in South Taihang Mountains (sublineage L4). However, sublineage L7 had a wide distribution across the upstream of Yellow River and Jialing River. Low levels of diversity indices were found in each sublineage ($\pi < 0.005$, Supplementary Table S5) and in most populations (Table 1). Particularly, the very low genetic diversity ($\pi < 0.001$, $Hd < 0.500$) was found in populations YYJ and YYS, distributed in the Luohe River system. Tajima's *D* and Fu's *Fs* were not significant in any sublineages, except for sublineage L7 ($p < 0.01$), as well as for all combined populations (Supplementary Table S5). AMOVA revealed that only 39.37% of genetic variation was explained by intra-population variation, while 60.63% ($p < 0.0000$) explained variation among populations (Table 2). Furthermore, a high differentiation at the group level was found in the present study, and 56.25% ($p < 0.0000$) explained variation among groups (Table 2). All these above results clearly identified a distinct hierarchical phylogeographical structure for the *T. robusta* in the studied area.

3.2 Demographic history and gene flow

The BSP detected relatively stable population sizes throughout the last 0.22 Mya for most sublineages (L1-L6) (data were not shown). However, the population size of sublineage L7 appeared to have increased constantly since 0.18 Mya (Supplementary Figure S1). Furthermore, the overall population size of the species appeared to have rapidly increased constantly since 0.21 Mya (Supplementary Figure S2). This suggested that only sublineage L7 in Northeastern QTP and the combined populations had expanded historically, coinciding with the result of the neutrality test (Supplementary Table S5).

Estimates of gene flow calculated with MIGRATE-n indicated low to high levels of historical gene flow between sublineages (Supplementary Table S6). The effective number of migrants entering and leaving each sublineage per generation ranged from $M = 0.29$ (L5→L4) to 15.44 (L5→L7).

TABLE 2 Analysis of molecular variance partitioning the genetic variance within and among populations, and molecular variance results comparing genetic variation among four groups (HS1-4 and TH1-4 as Yellow River system in Northeastern QTP; JLJ1-6 as Yangtze River system in Northeastern QTP; LBW, LSH, LSM, LNS, YYJ, YYS, XA, LCB, LCC, and SXY as Yellow River system in Qinling Mountains; other four populations as River systems in Southern Taihang Mountains).

Source of variation	Df	Sum of squares	Variance components	Percentage of variation with *p* value
Among populations	3	1285.646	7.38772 Va	60.63
Within populations	230	1103.277	4.79686 Vb	39.37
Total	233	2,388.923	12.18457	
Among four groups	3	1285.646	6.84779 Va	56.25
Among populations within groups	23	610.985	2.94756 Vb	24.21
Within populations	207	492.292	2.37822 Vc	19.54
Total	233	2388.923	12.17358	

The highest unidirectional estimates of gene flow were found towards the sublineage L7 (M = 37.41) and the lowest was in sublineage L4 (M = 2.04), while the highest gene flow out of a sublineage was from L5 (M = 18.14).

3.3 Divergence time estimation and ancestral range reconstruction

The time-calibrated molecular clock analyses dated the TMRCA of *T. robusta* and *T. siluroides* at approximately 8.51 Mya (95% highest posterior density HPD: 10.62–6.61 Mya) (Supplementary Figure S3). The divergence between sublineage L1 (Yihe River) and sublineage L2 (Luohe River) distributed in Eastern Qinling Mountains was at approximately 0.70 Mya, and both sublineages diverged from sublineages L5-7 in Northeastern QTP at approximately 2.17 Mya (HPD: 2.84–1.58 Mya). Sublineage L3 in Eastern Qinling Mountains split from L4 in South Taihang Mountains at approximately 0.61 Mya, and both sublineages diverged from L5-7 at approximately 1.00 Mya. Sublineage L5 upstream of the Yellow River diverged from L6 in Jialing River at approximately 0.65 Mya, and both diverged from sublineage L7 at approximately 0.87 Mya.

Ancestral range reconstruction showed that TMRCA of *T. robusta* was mainly distributed in the Yellow River basin (node 61, AC: 45.87%, Supplementary Figure S4). The Clade I most likely originated in the Eastern Qinling Mountains, while the Clade II originated in Northeastern QTP, indicating a vicariant event early in the history of *T. robusta*.

3.4 Species distribution modeling

Maxent modeling accurately predicted a current range similar to that known for *T. robusta* in China with little variance (Figures 2A,B), with AUC values of 0.967 (SD = 0.020; training AUC range: 0.972–0.980, test AUC range: 0.925–0.990). The temperature seasonality (bio4; 49.9%), Precipitation Seasonality (bio15; 24.4%), were the largest contributors to the model contributing 74.3%, however, Temperature Annual Range (bio7) and bio4 were the highest permutation importance (31.8%, 27.1%, respectively) as supported by jackknifing.

When the model used current conditions to predict suitable habitat for *T. robusta* during LGM, the main areas of suitable habitat with high probability for this species were as follows: the whole Qinling Mountains, Northeastern QTP, and Southern Taihang Mountains (Figure 2C). However, during LIG, suitable habitats reflected that of LIG were very different from LGM and the present distribution, with a greater concentrtion of suitable habitats in the Qinling Mountains and Northeastern QTP (Figure 2D).

4 Discussion

As a typical montane species, *T. robusta* is mainly distributed in Northeastern QTP and Qinling Mountains, providing an interesting model to explore the impact of paleoclimate oscillation and geologic events on diversification patterns. It was not uncommon that species restricted to high elevation mountain habitats often show high levels of intra-species divergence and even unique patterns of genetic structure (Pauls et al., 2006; Hou et al., 2014; de Oliveira et al., 2021). We found seven distinct sublineages of *T. robusta* with no shared haplotype among different mountain rivers, even in two close mountain rivers (Yihe River and Luohe River) in Qinling Mountains. However, we found two shared haplotypes (h33 and h62, Figure 1) distributed upstream of the Yellow River and Jialing River, which are divided by watershed in Northeastern QTP. The present study showed a novel phylogeographic pattern and evolution history for *T. robusta* compared with previous studies on this species. The interspecies

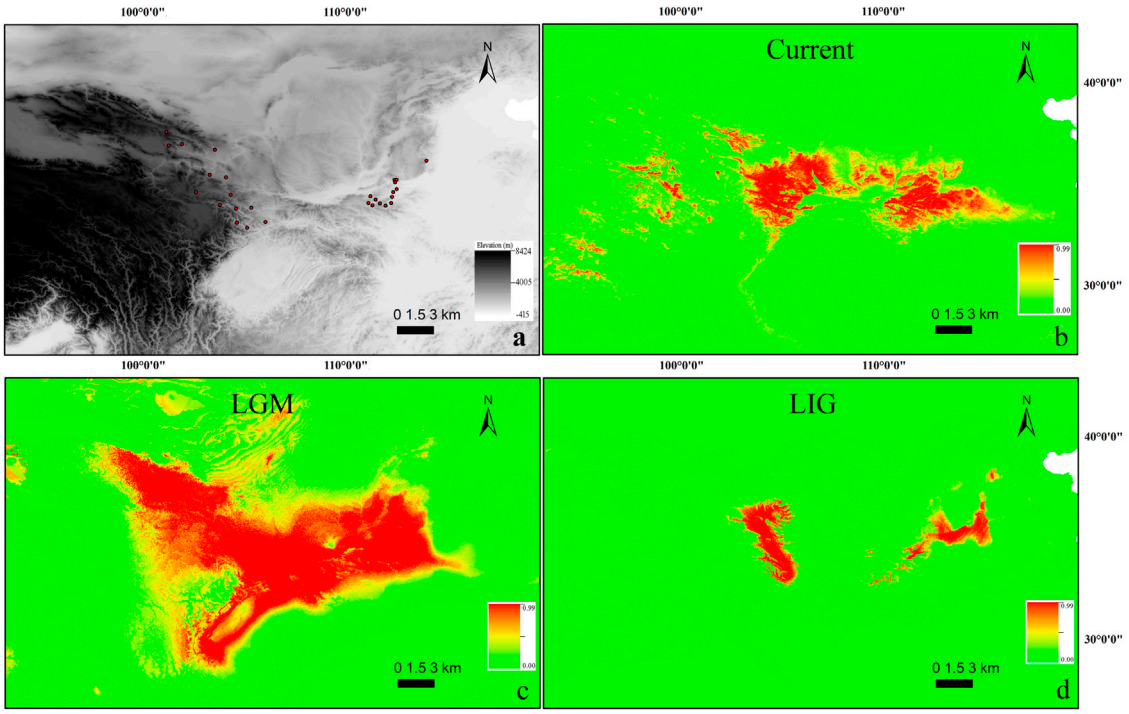


FIGURE 2
(A) Occurrence points used for building contemporary SDM for *T. robusta* (*n* = 29). The scale bar depicts altitude values. **(B–D)**: Species distribution modeling (SDM) for *T. robusta* during the current **(B)**, LGM **(C)**, and LIG **(D)** periods. Scale bars depict suitability values.

hybridization and introgressions in the *T. robusta* complex have distorted the molecular systematics and restricted speciation in the *T. robusta* complex (Feng et al., 2019). However, our study emphasized the historical gene flow among populations of *T. robusta*, resulting from the repetition of the glacial period during the Pleistocene, which played a crucial role in the phylogeographic pattern of *T. robusta*. Furthermore, our results showed that TMRCA of *T. robusta* differentiated into two clades (Figure 1) caused by a vicariant event early in the history of *T. robusta*, and Clade I most likely originated in the Eastern Qinling Mountains. While it was speculated that *T. robusta* in Qinling Mountains was mostly derived from the upstream of the Yellow River (Wu et al., 2020).

Our results revealed that *T. robusta* experienced Pleistocene rapid radiation. Divergence time among the seven sublineages was estimated at 2.17–0.61 Mya. Previous studies on plants and animals suggested mountain building as a predominant driver in triggering biological evolution (Hoorn et al., 2010; Smith et al., 2014; Simões et al., 2016), especially for the diversification of the species in QTP and Qinling Mountains (Che et al., 2010; Yan et al., 2010; Xu et al., 2018). However, there were few studies on the intra-species diversification between the two mountain lineages. In this study, we recovered the *T. robusta* in the investigated region of the present study with two main clades separated by the Qinling Mountains (Figure 1, Supplementary Figure S3). This mountain range extends more than 1,500 km from West to East in Central China with an average elevation between 1,500–3,000 m, representing an important geological and geographical boundary (Yan et al., 2010; Xu et al., 2018). The DIVALIKE + *j*-based analyses estimated allopatric speciation with the two sister clades inheriting each half of the ancestral range (Clade I in Eastern Qinling Mountains and Clade II in Northeastern QTP) (Supplementary Figure S4). The divergence time between clade I and clade II was estimated at approximately 2.17 Mya, relating to the further rapid uplift of the Qinling Mountains that occurred between 2.4 and 1.2 Mya (Yuan et al., 2012), and the interglacial period between 2.5 and 1.8 Mya (Xing, 1989; Jing and Liu, 1999). The retraction of cold-adapted species restricted to higher elevations during interglacial periods may have been another crucial element for producing a phylogeographic pattern for montane species (de Oliveira et al., 2021). The vicariance event likely had repeatedly happened during Pleistocene interglacials in Northeastern QTP and Qinling Mountains. Feng et al. (2019) mainly investigated the molecular systematics of the *T. robusta* complex distributed in Northeastern QTP. They speculated that there existed hybridization and introgressions between sympatric species according to the mitonuclear discordances, such as the hybridization and introgressions between *T. minxianensis* and *T. robusta*, and between *T. pappenheimi* and *T. siluroides*. Wu

et al. (2020) studied the molecular phylogeny and divergence of *Triplophysa* species distributed in the Qinling Mountains and Southern Taihang Mountains. However, the phylogeography pattern of *T. robusta* is still unknown according to the two previous studies. Fortunately, they supply a comprehensive distribution for the *T. robusta* in China, making a good opportunity for us to investigate the phylogeography of *T. robusta* across Northeastern QTP and the Qinling Mountains. The SDM suggested increases in habitat suitability for *T. robusta* during LGM (Figure 2C), but contraction during LIG (Figure 2D), especially in Qinling Mountains. Unidirectional migration was found in *T. robusta*, with middle to high level of *M* (>1.00) from sublineages (L1-L4) in Eastern Qinling Mountains to that (L5 and L7) in Northeastern QTP (Supplementary Table S6). This scenario of historical gene flow from east to west supported the existence of historical corridors for the dispersion of high-elevation species during periods of climate cooling (Antonelli et al., 2010; de Oliveira et al., 2021). However, we found a low level of migration rate in the opposite direction (a West-East pattern), which was uncommon in freshwater species found in China during the same or much older period, such as the frogs, snails, and fish (Yan et al., 2013; Gu et al., 2019; Gu et al., 2022). The authors suggested the role of Pleistocene interglacials dispersal in influencing the species diversification.

In the present study, the divergence estimated for the seven sublineages showed rapid radiation during 0.87–0.61 Mya (Supplementary Figure S3), exactly coinciding with the Pleistocene interglacial stage (0.90–0.40 Mya) in China. It was reported that the main uplift events of the Qinling Mountains with an elevation of about 3,000 m (Zhu, 1989) had been completed before the Early Pleistocene (1.67–1.45 Mya). The orogenic belt of the Taihang Mountains, with an average elevation of 1,500–2000 m, rose sharply to approximately 1,400 m at the Miocene/Pliocene boundary, and reached their current elevation at the end of the Pliocene (Gong, 2010). Many speciation events and intraspecific differentiation might have been driven by the massive uplifting of the Taihang Mountains and Qinling Mountains from the late Miocene to the Pliocene. Species such as the freshwater *Gammarus* with low dispersal ability showed a congruent pattern with geographical vicariance caused by the uplift of the Lüliang and Taihang Mountains (Hou et al., 2014). The uplift of the Qinling Mountains during 23–2.6 Ma as a vicariance event had driven the speciation of *Sinothela* (Xu et al., 2018). However, the recent Mid-Pleistocene diversifications in different sublineages of *T. robusta* were far postdated to the uplift of the two mountains. Therefore, the retractions during inter-glacial periods found in the present study were more likely to play a key role in interrupting gene flow for cold-adapted species. It was likely a good candidate explanation of the Mid Pleistocene rapid diversification for this species.

Strangely, sublineage L7 in Northeastern QTP had a wide distribution across the upstream of Jialing River and Yellow River, in which both possessed their own sublineages (L5 and L6, respectively). The nested genetic structure of *T. robusta* in Northeastern QTP suggested an exchange between the Yellow River and the Jialing River. The high level of historical gene flow from Jialing River to Yellow River supported the existence of historical corridors for the dispersal of *T. robusta* (Supplementary Table S6) and coincided with the speculation that the upstream of Yellow River was once connected to the Jialing River. It was supposed that the continual tectonic activities, such as earthquakes over the past million years in the Northeastern QTP likely led to extensive gene flow (Feng et al., 2019). Furthermore, during glacial periods, ephemeral rivers and periglacial lakes could arise, likely to provide further opportunities for dispersal and interactions between populations of *T. robusta*. The elevation shift model stated that populations of cold-adapted species inhabiting high elevations expanded to lowland areas during periods of climate cooling, increasing spatial connectivity and gene flow (Galbreath et al., 2009; Newman and Austin, 2015; de Oliveira et al., 2021).

Previous studies on the *T. robusta* (Feng et al., 2019; Wu et al., 2020) have little practical and instructive value on the conservation strategies of wildly distributed species of *Triplophysa*. The present study had an important implication for the conservation of the cold-adapted species restricted to higher elevations. The diversification pattern found in the *T. robusta* can provide guidelines for effective conservation measures for *Triplophysa*, which might be negatively impacted by current climate change caused by humans. Montane endemic species were thought to be particularly vulnerable to the effects of rapid climate changes, as lots of montane habitats may be lost or fragmented due to climate warming (Moritz et al., 2008). Bonatelli et al. (2014) and Fernandes et al. (2014) believed that climate change in the next few decades might accelerate the declines and local extinctions of high elevation species due to the reduction of suitable habitats. The low genetic diversity found in most sublineages and populations of *T. robusta* indicated a serious challenge in adaptive ability (O'Brien, 1994; Bazin et al., 2006). Our results suggested that the temperature was one of the most important environmental factors influencing *T. robusta* distribution range, which experienced a considerable retraction as historical climatic fluctuations in Northeastern QTP and Qinling Mountains in China. In this context, the studied species *T. robusta*, as well as other *Triplophysa* distributed in the QTP and adjacent regions would be negatively impacted by climate warming. A burning question in future conservation projection on cold-adapted fish species endemic to higher elevations is how to tackle global warming for them. Furthermore, mid to low levels of genetic diversities were found for all populations except for JYT (Table 1). While *T. robusta* is classed as Least Concern (LC) in China, the current conservation strategy is carried out under the assumption of only one species and one gene pool. Our results demonstrated that

different protection strategies for *T. robusta* should be conducted in different mountains and river systems: two protection units for *T. robusta* in Northeastern QTP, one upstream of the Yellow River basin, and the other one upstream of the Jialing River; three protection units in Qinling Mountains, one in Yihe River, one in Luohe River and another in Dongjian River; and only one protection unit in the Southern Taihang Mountains. Especially for two populations (YYJ and YYS) in a couple of mountain streams in Luohe River, with very low nucleotide diversity and haplotype diversity, forbidden fishing zone and habitat environment restoration should be implemented to strengthen protection for them.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found in the article/Supplementary Material.

Ethics statement

The animal study was reviewed and approved by the Hunan Normal University and Henan Normal University.

Author contributions

QG conceived and designed the study. HZ and YS analyzed the data. HW, HZ, SL, ZS, and CY carried out tissue dissection, DNA isolation, sample preparation for sequencing and collecting the mtDNA data from GenBank. QG and HZ drafted the manuscript, and MW and PC revised the language. All authors read and approved the final manuscript.

Funding

This study was supported by the Natural Science Foundation of Hunan (Grant No. 2021JJ30442, 2020JJ5357), and the National Natural Science Foundation of China (No. 31601851, 32002372).

Acknowledgments

We greatly appreciate the reviewers for their very helpful comments on previous versions of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.955382/full#supplementary-material

References

Anderson, R. P., and Raza, A. (2010). 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.* 37, 1378–1393. doi:10.1111/j.1365-2699.2010.02290.x

Bazin, E., Glémin, S., and Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572. doi:10.1126/science.1122033

Bonatelli, I. A. S., Perez, M. F., Peterson, A. T., Taylor, N. P., Zappi, D. C., Machado, M. C., et al. (2014). Interglacial microrefugia and diversification of a cactus species complex: Phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Mol. Ecol.* 23, 3044–3063. doi:10.1111/mec.12780

Che, J., Zhou, W. W., Hu, J. S., Yan, F., Papenfuss, T. J., Wake, D. B., et al. (2010). Spiny frogs (paini) illuminate the history of the himalayan region and southeast asia. *Proc. Natl. Acad. Sci. U. S. A.* 107, 13765–13770. doi:10.1073/pnas.1008415107

Clark, M. K., House, M. A., Royden, L. H., Whipple, K. X., Burchfiel, B. C., Zhang, X., et al. (2005). Late cenozoic uplift of southeastern tibet. *Geol.* 33, 525–528. doi:10.1130/g21265.1

de Oliveira, F. R., Gehara, M., Solé, M., Lyra, M., Haddad, C. F. B., Silva, D. P., et al. (2021). Quaternary climatic fluctuations influence the demographic history of two species of sky-island endemic amphibians in the Neotropics. *Mol. Phylogenet. Evol.* 160, 107113–113. doi:10.1016/j.ympev.2021.107113

Dong, Y., Zhang, G., Neubauer, F., Liu, X., Genser, J., and Hauzenberger, C. (2011). Tectonic evolution of the Qinling orogen, China: Review and synthesis. *J. Asian Earth Sci.* 41, 213–237. doi:10.1016/j.jseaes.2011.03.002

Feng, C. G., Zhou, W. W., Tang, Y. T., Gao, Y., Chen, J. M., Tong, C., et al. (2019). Molecular systematics of the *Triplophysa robusta* (cobitoidea) complex: Extensive gene flow in a depauperate lineage. *Mol. Phylogenet. Evol.* 132, 275–283. doi:10.1016/j.ympev.2018.12.009

Fernandes, G. W., Barbosa, N. P. U., Negreiros, D., and Paglia, A. P. (2014). Challenges for the conservation of vanishing megadiverse rupestrian grasslands. *Natureza Conservacao* 12, 162–165. doi:10.1016/j.ncon.2014.08.003

Galbreath, K. E., Hafner, D. J., and Zamudio, K. R. (2009). When cold is better: Climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American Pika, *Ochotona princeps*). *Evolution* 63, 2848–2863. doi:10.1111/j.1558-5646.2009.00803.x

Gong, M. (2010). *Uplifting process of southern Taihang mountian in cenozoic*. PhD Thesis. Beijing, China: Chinese Academy of Geological Science.

Gu, Q. H., Husemann, M., Wu, H. H., Dong, J., Zhou, C. J., Wang, X. F., et al. (2019). Phylogeography of *bellamya* (Mollusca: Gastropoda: Viviparidae) snails on different continents: Contrasting patterns of diversification in China and east africa. *BMC Evol. Biol.* 19, 82. doi:10.1186/s12862-019-1397-0

Gu, Q. H., Wang, S., Zhong, H., Yuan, H., Yang, J. L., Yang, C. H., et al. (2022a). Phylogeographic relationships and the evolutionary history of the *Carassius auratus* complex with a newly born homodiploid raw fish (2nNCRC). *BMC Genomics* 23, 242. doi:10.1186/s12864-022-08468-x

Gu, Q., Zhong, H., Sun, Y., Yuan, H., Li, S., Shen, Z., et al. (2022b). Reanalysis on phylogeographic pattern of sharpbelly *Hemiculter leucisculus* (cyprinidae: Cultrinae) in China: A review and the implications for conservation. *Front. Ecol. Evol.* 10, 865089. doi:10.3389/fevo.2022.865089

Hasumi, H., and Emori, S. (2004). *K-1 coupled model (MIROC) description. K-1 Tech Rep. p 34*. Available at: http://ccsr.aori.u-tokyo.ac.jp/~hasumi/miroc_description.pdf (Accessed at Sep 29, 2014).

Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., et al. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330, 927–931. doi:10.1126/science.1194585

Hou, Z., Li, J. B., and Li, S. Q. (2014). Diversification of low dispersal crustaceans through mountain uplift: A case study of *Gammarus* (Amphipoda: Gammaridae) with descriptions of four novel species. *Zool. J. Linn. Soc.* 170, 591–633. doi:10.1111/zoj.12119

Huang, Z., Yu, F., Gong, H., Song, Y., Zeng, Z., and Zhang, Q. (2017). Phylogeographical structure and demographic expansion in the endemic alpine stream salamander (hynobiidae: *Batrachuperus*) of the qinling mountains. *Sci. Rep.* 7, 1871. doi:10.1038/s41598-017-01799-w

Husemann, M., Guzman, N. V., Danley, P. D., Cigliano, M. M., and Confalonieri, V. A. (2013). Biogeography of *Trimerotropis pallidipennis* (acrididae: Oedipodinae): Deep divergence across the americas. *J. Biogeogr.* 40, 261–273. doi:10.1111/jbi.12007

Jing, C. R., and Liu, H. P. (1999). On the glacial and Interglacial stages in Quaternary of China. *J. Chengdu Univ. Technol.* 26, 97–100.

Kong, H. H., Condamine, F. L., Yang, L. H., Harris, A. J., Feng, C., Wen, F., et al. (2022). Phylogenomic and macroevolutionary evidence for an explosive radiation of a plant genus in the Miocene. *Syst. Biol.* 71 (3), 589–609. doi:10.1093/sysbio/syab068

Li, J. X., Wang, Y., Jin, H. F., Li, W. J., Yan, C. C., Yan, P. F., et al. (2017). Identification of *Triplophysa* species from the Qinghai-Tibetan Plateau (QTP) and its adjacent regions through DNA barcodes. *Gene* 605, 12–19. doi:10.1016/j.gene.2016.11.045

Liu, J., Zhang, P., Lease, R. O., Zheng, D., Wan, J., Wang, W., et al. (2013). Eocene onset and late Miocene acceleration of Cenozoic intracontinental extension in the North Qinling range-Weihe graben: Insights from apatite fission track thermochronology. *Tectonophysics* 584, 281–296. doi:10.1016/j.tecto.2012.01.025

Meng, Q. (2017). Origin of the qinling mountains (in Chinese). *Sci. Sin. Terrae* 47, 412–420.

Moritz, C., Patton, J. L., Conroy, C. J., Parra, J. L., White, G. C., and Beissinger, S. R. (2008). Impact of a century of climate change on small-mammal communities in Yosemite national park, USA. *Science* 322, 261–264. doi:10.1126/science.1163428

Newman, C. E., and Austin, C. C. (2015). Thriving in the cold: Glacial expansion and postglacial contraction of a temperate terrestrial salamander (*Plethodon serratus*). *PLoS ONE* 10, e0130131. doi:10.1371/journal.pone.0130131

O'Brien, S. J. (1994). A role for molecular genetics in biological conservation. *Proc. Natl. Acad. Sci. U. S. A.* 91 (13), 5748–5755. doi:10.1073/pnas.91.13.5748

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., and Hu, A. (2006). Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311, 1751–1753. doi:10.1126/science.1120808

Pauls, S. U., Lumbsch, T., and Haase, P. (2006). Phylogeography of the montane caddisfly *Drusus discolor*: Evidence for multiple refugia and periglacial survival. *Mol. Ecol.* 15, 2153–2169. doi:10.1111/j.1365-294X.2006.02916.x

Perdices, A., Vasil'Eva, E., and Vasil'Ev, V. (2014). From asia to europe across siberia: Phylogeography of the siberian spined loach (teleostei, Cobitidae). *Zool. Scr.* 44 (1), 29–40. doi:10.1111/zsc.12085

Rice, W. R., and Hostert, E. E. (1993). Laboratory experiments on speciation: What have we learned in 40 years? *Evolution* 47, 1637–1653. doi:10.1111/j.1558-5646.1993.tb01257.x

Schramm, F. D., Valdez-Mondragón, A., and Prendini, L. (2021). Volcanism and palaeoclimate change drive diversification of the world's largest whip spider (Amblypygi). *Mol. Ecol.* 30, 2872–2890. doi:10.1111/mec.15924

Simões, M., Breitkreuz, L., Alvarado, M., Baca, S., Cooper, J. C., Heins, L., et al. (2016). The evolving theory of evolutionary radiations. *Trends Ecol. Evol.* 31, 27–34. doi:10.1016/j.tree.2015.10.007

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., et al. (2014). The drivers of tropical speciation. *Nature* 515, 406–409. doi:10.1038/nature13687

Sota, T., and Vogler, A. P. (2001). Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *ohomopterus*. *Syst. Biol.* 50, 39–59. doi:10.1080/10635150119427

Wang, B., Jiang, J., Xie, F., and Li, C. (2012). Postglacial colonization of the qinling mountains: Phylogeography of the swelled vent frog (*Feirana quadranus*). *PLoS ONE* 7, e41579. doi:10.1371/journal.pone.0041579

Wang, X. G., Ma, T., Yun, Y., Li, Y., Zhang, J., Liang, H. Y., et al. (2021a). Distinct phylogeographic structures and evolutionary histories of wild medicinal *Salvia miltiorrhiza* Bunge plant populations in the mountains of Central China. *Genet. Resour. Crop Evol.* 68, 3309–3318. doi:10.1007/s10722-021-01189-2

Wang, X. Z., Gan, X. N., Li, J. P., Chen, Y. Y., and He, S. P. (2016). Cyprininae phylogeny revealed independent origins of the Tibetan Plateau endemic polyploid cyprinids and their diversifications related to the Neogene uplift of the plateau. *Sci. China. Life Sci.* 59, 1149–1165. doi:10.1007/s11427-016-0007-7

Wang, Y., Feijó, A., Cheng, J., Xia, L., Wen, Z. X., Ge, D. Y., et al. (2021b). Ring distribution patterns—diversification or speciation? Comparative phylogeography of two small mammals in the mountains surrounding the sichuan basin. *Mol. Ecol.* 30, 2641–2658. doi:10.1111/mec.15913

Wu, H. H., Gu, Q. H., Zhou, C. J., Tang, Y. T., Husemann, M., Meng, X. L., et al. (2020). Molecular phylogeny and biogeography of *Triplophysa* stone loaches in the Central Chinese Mountains. *Biol. J. Linn. Soc. Lond.* 130 (3), 563–577. doi:10.1093/biolinnean/blaa060

Xing, L. S. (1989). Magnetostratigraphic age of Quaternary glaciations in the Lushan area. *Bull. Inst. Geomechanics Cags* 13, 71–79. (in Chinese).

Xu, X., Kuntner, M., Liu, F. X., Chen, J., and Li, D. (2018). Formation of rivers and mountains drives diversification of primitively segmented spiders in continental East Asia. *J. Biogeogr.* 45, 2080–2091. doi:10.1111/jbi.13403

Yan, J., Wang, Q., Chang, Q., Ji, X., and Zhou, K. (2010). The divergence of two independent lineages of an endemic Chinese gecko, *Gekko swinhonis*, launched by the Qinling orogenic belt. *Mol. Ecol.* 19, 2490–2500. doi:10.1111/j.1365-294X.2010.04660.x

Yuan, J. H., Cheng, F. Y., and Zhou, S. L. (2012). Genetic structure of the tree peony (*Paeonia rockii*) and the Qinling Mountains as a geographic barrier driving the fragmentation of a large population. *PLoS ONE* 7, e34955. doi:10.1371/journal.pone.0034955

Zhang, R., Jiang, D. B., Liu, X. D., and Tian, Z. P. (2012). Modeling the climate effects of different subregional uplifts within the Himalaya-Tibetan Plateau on Asian summer monsoon evolution. *Chin. Sci. Bull.* 57, 4617–4626. doi:10.1007/s11434-012-5284-y

Zhu, Z. Y. (1989). The formation of river terraces and evolution of drainage system in the middle Yellow River. *Acta Geogr. Sin.* 44, 429–440. 10.11821/xb198904006.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fgene.2022.963341

OPEN ACCESS

EDITED BY
Narongrit Muangmai,
Kasetsart University, Thailand

REVIEWED BY
Jianping Xu,
McMaster University, Canada
Clint Rhode,
Stellenbosch University, South Africa
Sirawut Klinbunga,
National Science and Technology Development Agency, Thailand

*CORRESPONDENCE
A. Gilles,
andre.gilles@univ-amu.fr

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 07 June 2022
ACCEPTED 02 September 2022
PUBLISHED 23 September 2022

CITATION
Gilles A, Thevenin Y, Dione F, Martin J-F, Barascud B, Chappaz R and Pech N (2022), Breaking the reproductive barrier of divergent species to explore the genomic landscape. *Front. Genet.* 13:963341. doi: 10.3389/fgene.2022.963341

Breaking the reproductive barrier of divergent species to explore the genomic landscape

A. Gilles[1]*, Y. Thevenin[1], F. Dione[1], J.-F. Martin[2], B. Barascud[1], R. Chappaz[1] and N. Pech[1]

[1]Aix Marseille University, INRAE, UMR 1467 RECOVER, Centre Saint-Charles, Marseille, France, [2]CBGP, Montpellier SupAgro, INRA, CIRAD, IRD, Université Montpellier, Montpellier, France

Background: Climate change will have significant consequences for species. Species range shifts induce the emergence of new hybrid zones or the spatial displacement of pre-existing ones. These hybrid zones may become more porous as alleles are passed from one species to another. Currently, hybridization between highly divergent species living in sympatry seems extremely limited. Indeed, this phenomenon involves breaking two barriers. The first is the pre-mating barrier, related to the reproductive phenology of the two species. The second is the post-zygotic barrier, related to the genetic divergence between these species. Here, we were interested in identifying new hybridization patterns and potential implications, especially in the context of environmental modifications.

Methods: We sampled *Telestes souffia* and *Parachondrostoma toxostoma* wild specimens from different locations across France and genotyped them for SNP markers. We identified discriminant loci using F1-hybrid specimens and parental species and performed principal component analysis and Bayesian model-based clustering to analyze phylogenetic information. Furthermore, we assessed deviation in allele frequency from F1 to F2 and for Hardy–Weinberg equilibrium for F2 and assessed gene function associated with two F2 cohorts.

Results: We demonstrate that by breaking the ecological barrier, massive introgressive hybridization is possible between two endemic lineages of Cyprinidae belonging to two distinct genera. For both cohorts studied (=2 cm and >2 cm), a large majority of loci (>88%) presented no deviation in allele frequency and no departure from the Hardy–Weinberg equilibrium. For individuals beyond the 2 cm stage, two phenomena were observed. The first was an allelic imbalance in favor of *P. toxostoma*, for some genomic regions, with genes involved in developmental regulatory processes, cytoskeletal organization, and chromosome organization. The second was an excess of heterozygous loci coupled with an equilibrium of allelic frequencies for genes involved in immune response and kidney/liver development. Moreover, the 2 cm-sized specimens with high mortality yielded a particular genomic signature.

Conclusion: Our study displayed important results for understanding the early stages of hybridization between divergent lineages and predicting the

emergence of future hybrid zones in the wild. Moreover, this hybridization generates a wide spectrum of hybrids that are a potential source of important evolutionary novelties.

KEYWORDS

speciation, SNP, RAD sequencing, hybridization, genomic introgression, transcriptome

1 Introduction

As early as 1948, Anderson (1948) hypothesized that the impact of anthropogenic changes would affect habitats in such a way that hybridization events would be more frequent and hybrid survival would be greater. Although some authors (Lewontin and Birch, 1966) have described hybridization as a source of variation allowing adaptation to new environments, a predominantly negative view of hybridization has long prevailed, illustrated by terms such as 'extinction by hybridization,' 'genetic pollution,' or 'evolutionary dead end' (Rhymer and Simberloff, 1996). However, hybridization can generate evolutionary novelties, such as transgressive segregation at a much faster rate than intra-specific adaptation, to transfer new alleles (adaptive introgression of alleles), allowing, for example, the evolutionary rescue of a lineage living in extreme pollution conditions (Oziolor et al., 2019), or new species induction (speciation by hybridization) in new environments, either by allopolyploidy (when the ploidy level of the hybrid populations is equal to the sum of the chromosomes of each parent species) or by homopolyploidy (when the ploidy level of the hybrid populations and that of the parent species do not change). This paradigm shift in hybridization is due to increased research in the field. Indeed, our knowledge of hybridization in the wild has increased considerably and has led to a better understanding of the mechanisms involved and its evolutionary potential (Runemark et al., 2019).

However, the ability of two species to hybridize and the fate of hybrid individuals is difficult to predict. Indeed, it is unclear how to predict future hybridization behaviors for species that do not currently hybridize. Furthermore, there may be genomic porosity between phylogenetically distant species that are hidden by current environmental conditions. Owens and Samuk (2020) modeled the effects of climate change on hybrid zones and showed that hybrid zones become increasingly porous as climate change-adapted alleles are passed from one species to another, and that this effect is stronger when climate change is greater. Other authors have proposed models to predict the incompatibility of hybrids with the parental species that generated them (Schumer et al., 2015).

There is growing recognition that climate change will reshape species assemblages and shift spatial, temporal, and behavioral reproductive barriers (Chunco, 2014). Climate change leads to shifts in species ranges that can result in the displacement of pre-existing hybrid areas, such as for the two chickadee species *Poecile atricapillus* and *Poecile carolinensis* (Taylor et al., 2014), or new hybrid areas for the two hare species *Lepus timidus* and *Lepus europaeus* (Jansson et al., 2007). Thus, climate change may alter not only the respective densities of species present in the environment but also the sex ratio within each species. In both cases, heterospecific crosses can occur and lead to the existence of hybrids. Under these conditions, some hybrid combinations could emerge in these new environments and be selected positively or negatively (exogenous selection). However, for this to happen, it is first necessary for the genomic compatibility between the species to be possible (endogenous selection) and, second, to overcome the barrier of F1 individuals either by backcrossed specimens or by F2-hybrids.

Many studies (Edmands, 2002) have investigated whether there is a correlation between the genetic divergence of parental species and their level of reproductive incompatibility (pre- or post-zygotic). In this case, the greater the phylogenetic divergence, the lower the genetic compatibility, and vice versa. Some authors (Orr and Turelli, 2001) have proposed the 'snowball' model to model the evolution of genetic incompatibility with genetic divergence or time and showed that the expected number of hybrid incompatibilities rose exactly with the square of time. This was empirically verified in two scientific papers, one on the genus *Drosophila* (Matute et al., 2010) and the other on the genus *Solanum* (Moyle and Nakazato, 2010). Moreover, this genetic incompatibility, which increases with genetic divergence, is accompanied by an increase in transgressive phenotypes due to divergent alleles from the two parental species (Stelkens and Seehausen, 2009; You and Xu, 2021), leading to non-additive effects by over-dominance, in addition to epistatic interactions, such as the modification of the expression of a gene by other genes (Stelkens and Seehausen, 2009). In the case of Cichlidae of the Haplochromine group, genetic divergence between species explains 52% of the variation in transgressive traits for F1-hybrids and 78% for F2-hybrids (Stelkens and Seehausen, 2009). Thus, hybridization between two different genera is rare, but such cases exist in some Teleost families, such as Cyprinidae (Scribner et al., 2000). The majority of these hybrids are the result of crossing between an endemic species and a closely related introduced genus, as in the case of *Chondrostoma nasus* and *Parachondrostoma toxostoma* resulting in fertile hybrids (Costedoat et al., 2007), or between two introduced species, as in the case of *Rutilus rutilus* and

Abramis brama (Hayden et al., 2014; Konopiński and Amirowicz, 2018) yielding mostly F1-hybrids. Finally, some rare cases involve a natural hybridization of two endemic species, as in *Achondrostoma oligolepis* and *Pseudochondrostoma duriense* (Pereira et al., 2014). As Canestrelli (Canestrelli et al., 2016) aptly pointed out: 'Not surprisingly, species of ancient divergence and with a long-lasting history of sympatry have had ample opportunity to evolve strong pre-mating reproductive barriers'.

Given the existence of hybridization phenomena, we were interested in identifying new ones, especially in the context of environmental modifications. Moreover, hybridization is a common phenomenon in fishes in general and cyprinids in particular (Smith 1992; Dowling and DeMarais, 1993). We, therefore, proposed the following question: To what extent is it possible to qualify and quantify the potential porosity of evolutionary barriers between two species that may be very divergent? In order to answer this question, we were interested in two endemic species of southern France, *Parachondrostoma toxostoma* (Vallot, 1837) or toxostome and *Telestes souffia* (Risso, 1826) or blageon. These two species live in sympatry and are phylogenetically distant with a time of divergence estimated at 15 million years, based on the molecular marker whole cytochrome b (Doadrio and Carmona, 2004; Perea et al., 2010). These two species have an extremely low natural hybridization rate since only two hybrid specimens have been detected in the wild (Dubut et al., 2010; Dubut et al., 2012), and such hybrids had been identified in the past, which led to the invalidation of the species *Chondrostoma rysela* (Günther 1868 in (Roule, 1925; Kottelat, 1997)). The rarity of these hybrids can be explained not only by very strong genetic incompatibility due to phylogenetic distance but also by the existence of ecological barriers. Indeed, *P. toxostoma* specimens reproduce (single spawn) essentially from April until the beginning of May in the range of water temperature from 11 to 13°C in the Buech tributary of Durance, France (Costedoat et al., 2007), whereas *T. souffia* specimens reproduce (single spawn) during the month of June, in water temperatures from 10 to 14°C. Warmer temperatures and low water levels in May–June will cause the shift of breeding season to the earlier end of March and beginning of May. The most impacted species will be the *T. souffia* because June would not present the appropriate conditions for reproduction. Moreover, the densities of the two species are reversed along an upstream–downstream thermal gradient (mostly *T. souffia* upstream, *P. toxostoma* downstream). For this study, we investigated 369 specimens represented by field-sampled (*T. souffia* and *P. toxostoma* 'pure' specimens) and laboratory-bred individuals (*P. toxostoma* 'pure' specimens, F1-hybrids, F2-hybrids, and backcrossed specimens), and 5,289 single nucleotide polymorphisms (SNPs) based on the *P. toxostoma* transcriptome.

2 Materials and methods

2.1 Presentation of the species and populations used

The *Telestes souffia* or blageon is a species of about 15 cm with a purplish stripe on the flanks, especially in males. It is protected but considered to be of least concern on the IUCN global red list (2008 assessment, https://www.iucnredlist.org/species/61397/12461824) and the red list of freshwater fishes of metropolitan France (2019). We used specimens belonging to the lineage *T. s. souffia*, which is specific to the Rhône and Mediterranean coastal river basins (from the Hérault River to the Var River) (Dubut et al., 2012). The *Parachondrostoma toxostoma* or toxostoma is a species with a tapered body, 15–25 cm long, olive-green in color, with light-colored sides showing silver reflections and a dark band that stands out, particularly during the breeding season. It is protected and is considered vulnerable on the IUCN world red list (2006 assessment, https://www.iucnredlist.org/species/4795/11095903) and near threatened on the red list of freshwater fishes of metropolitan France (2019). Its range extends from the southwest (the Adour–Garonne basin) to the southeast (the Rhône and Mediterranean coastal river basins) of France. Both species are diploid and have an identical number of homologous chromosomes (2n = 50).

2.2 Wild samples and specimen crosses

To calibrate SNP discriminant loci, we used wild specimens sampled in different locations (see Figure 1 and Table 1). For the *T. souffia* species, we selected four populations from the Rhône basin, especially in the Durance River: ARC_Ts (Archidiacre station on Durance River Lat. 44.475,719 Lon. 6.111,014), MAN_Ts (Manosque on Durance River Lat. 43.8,402,500 Lon. 5.8,506,111), BUE_Ts (Buech River Lat. 44.344,493 Lon. 5.770,268), and AIN_Ts (Ain River Lat. 46.264,383 Lon. 5.429,392), which were previously described (Dubut et al., 2012). For *P. toxostoma*, because of natural hybridization with *Chondrostoma nasus*, we selected three populations in parapatry from the Rhône basin, with no sign of hybridization—DOU_Pt (Doubs River Lat. 47.4,812,500 Lon. 6.8,360,556), SEP_Pt (Serre-Ponçon dam Lat. 44.536,424 Lon. 6.417,518), and AIN_Pt (Ain River Lat. 46.264,383 Lon. 5.429,392)—and two in allopatry—BER_Pt (Berre River Lat. 43.032721 Lon. 2.831,686) and ORB_Pt (Orbieu River Lat. 43.101,508 Lon. 2.623,265)—which were previously described (Costedoat et al., 2007; Corse et al., 2015).

In 2011, 14 F1-hybrids specimens were obtained by *in vitro* fertilization of 50 eggs and semen between one female T. souffia

TABLE 1 SNP characteristics mapping.

Level of SNP detection on *P. toxostoma* transcriptome	Loci number	Number of conserved loci	Number of SNPs	Groups	Tissue collection year	Number of individuals	Number of unique individuals	Number of sequencing replicates	Number of extraction replicates	Sequencing run	Number of reads
70% (within populations and between populations)	89,787	10,099 (11.25%)	56,283 (5,289 variable sites)	ARC_Ts	1996	3	3			run1	7,717,035
				MAN_Ts	1996	4	4			run1	8,073,878
				ORB_Pt	2007	5	5			run1	12,728,611
				AIN_Pt	2000	5	5			run1	11,393,464
				F1_Hy_R1	2015	19	14	5		run1	181,875,041
				F2_Hy	2021	213	204	4	5	run2	1,597,097,231
				BK_Hy	2015	15	10	1	4	run1	197,879,261
				AIN_Ts	1996	12	11	1		run1	46,468,500
				DOU_Pt	1996	5	5			run1	33,934,941
				SEP_Pt	1997	7	6	1		run1	24,326,823
				LAB_Pt	2021	33	33			run2	674,783,611
				F1_Hy_R2	2021	9			9	run2	42,338,828
				GEN_Pt	2011	5	5			run2	195,101,630
				BUE_Ts	2021	24	24			run2	176,614,316
				BER_Pt	2007	10	10			run2	40,999,443

P. toxostoma populations: AIN_Pt for Ain River, DOU_Pt for Doubs River, ORB_Pt for Orbieu River, BER_Pt for Berre River, GEN_Pt for genitors, LAB_Pt for laboratory specimens, SEP_Pt for Serre-Ponçon dam. *T. souffia* populations: AIN_Ts for Ain River, ARC_Ts for Archidiacre River, BUE_Ts for Buech River, MAN_Ts for Manosque on Durance River. F1_Hy_R1, first run of F1 laboratory specimens; F1_Hy_R2, second run of F1 laboratory specimens; F2_Hy. F2 specimens; BK_Hy, backcrossed laboratory specimens between F1 and *P. toxostoma* laboratory specimens.

10.3389/fgene.2022.963341

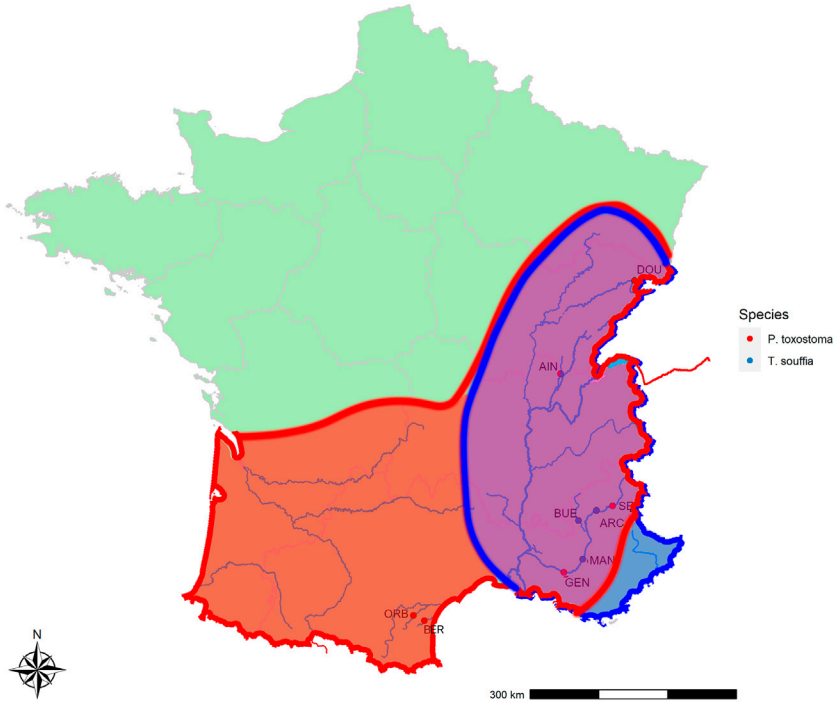


FIGURE 1
Species distribution and collection sites. In red *P. toxostoma* sites, in blue T. souffia sites.

(from the population BUE_Ts) and one male P. toxostoma (from the population GEN_Pt, inhabiting the Durance River). Free-swimming larvae were first housed in small aquaria (40 × 25 × 15 cm); after 4 months, the fish were transferred into larger aquaria for growth (100 × 60 × 40 cm). Fish were fed live *Artemia* nauplii, complemented with commercial fish food. The P. toxostoma laboratory stock (LAB_Pt) was obtained by crossing one female (population GEN_Pt from the Durance River at Pertuis) and two males (population GEN_Pt from the Durance River at Pertuis). In 2014, nine backcrosses were obtained by crossing one female P. toxostoma (3 years old) from the laboratory stock and three male F1-hybrids (3 years old). Furthermore, 10 F2-hybrids were obtained in 2014 and 194 in 2018 by crossing eight F1-hybrid specimens (3 and 7 years old) by crossing two females and six males (random mating from the F1 population).

2.3 Molecular protocol

In total, 369 individuals (including 12 DNA extraction replicates and 18 tissue replicates on specimens alive 4 years later, that constituted 8.13% of replicates) belonging to the two species and their hybrids (F1-hybrids, F2-hybrids, and backcross) were genotyped for SNP markers by first digesting genomic DNA with PstI restriction endonuclease as previously described (Brazier et al., 2021) (Supplementary Material) with the modifications described in Supplementary Figure S1.

We sequenced the mitochondrial DNA (1,140 bp of the cytochrome b) for the whole hybrid specimens, corresponding to only two haplotypes: *T. souffia* for F1-hybrids and F2-hybrids and *P. toxostoma* for backcross (Supplementary Data S1).

A sequencing Illumina Hiseq 2000 1 × 100 bp (Single End) run was generated with 5% Phi-X Spike-in by Cornell University (https://www.biotech.cornell.edu/facilities-brc). Then, two sequencing Illumina NovaSeq 2 × 150 bp (Pair End) runs were generated with 10% Phi-X Spike-in by NOVOGENE (https://en.novogene.com). The quality of the three sequencing runs was checked independently using FastQC, and the demultiplexing step was performed using the 'process_radtags' program implemented in the Stacks software (2.53) (Catchen et al., 2013; Rochette and Catchen, 2017). All the samples were analyzed in single-end and submitted to SRA (Sequence Read Archive, https://www.ncbi.nlm.nih.gov/sra, BioProject: PRJNA863842Submission ID: 30065804-30066172). The demultiplexing step was performed as follows: reads with an uncalled base were removed, a quality Phred score = 33 was used, and reads were truncated to 80 bp (Min-Max). The corresponding command was:

2.3.1 Process_radtag-f/raw_fastq-o/sample/-b/pst_barcodes.csv–inline_null–renz_1 pstI -r -c -q -t 80 –len_limit 80

We conserved 99% of reads per individual. We used the genome mapping option of Stacks software (2.53) on the *P. toxostoma* transcriptome (Ungaro et al., 2017). The transcriptome displayed orthologous ENSDARG annotation (Ensembl, https://www.ensembl.org, *Danio rerio*, gene) because it corresponded to a model system belonging to the same family, Cyprinidae. We used the genome nomenclature (ENSDARG) rather than transcriptome nomenclature (ENSDART) because we used the chromosome and gene position of *D. rerio* for mapping the discriminant SNPs between the two species *P. toxostoma* and *T. souffia*. Then, we used the 'ref_map.pl' program in the Stacks software (2.53). The corresponding command was:

2.3.2 Ref_map.pl -T28–popmap/popmap.csv -o/stacks/--samples/sample/

The program consisted of four principal steps (Paris et al., 2017): 1) Short read alignments were performed using BWA-MEM (Li and Durbin, 2009; Li, 2013) on the *P. toxostoma* transcriptomes. 2) The output data in SAM file format was sorted and converted to BAM file format using the software SAMtools (Li et al., 2009). 3) Then, SNP scoring and genotype identification was performed on reads alignment for each sample using the 'Gstacks' program of the Stacks software. 4) Finally, we conserved only the loci shared between 70% of the individuals within and between groups using the Populations software (Catchen et al., 2013; Rochette and Catchen, 2017).

2.4 Analysis of single-nucleotide polymorphism data

Selecting useful SNPs from Restriction-site Associated DNA-sequencing (RAD-sequencing) constitutes an important step in data analysis (Díaz-Arce and Rodríguez-Ezpeleta, 2019). Thus, we developed an approach to identify and select discriminant loci using F1-hybrid specimens and parental species (using different populations). The initial dataset consisted of 369 individuals and 5,289 SNPs. The individuals corresponded to five groups: wild populations of *P. toxostoma* (70 individuals from seven populations), wild populations of *T. souffia* (43 individuals from four populations), and farmed individuals—F1 (28 individuals), F2 (213 individuals), and backcrossed (15 individuals).

We defined the Euclidean distance between two specimens or replicates as the square root of the sum of squares of genotypic values (0 for homozygous *P. toxostoma*, 1 for heterozygotes, and 2 for homozygous *T. souffia*) based on the whole loci. The distribution of the 67,896 pairwise distances showed a multimodal distribution with three modes. We hypothesized that the first mode, corresponding to the 34 lower distances, included only the DNA and/or extraction replicates. Considering this criterion, we removed 18 specimens: twelve corresponded to DNA replicates (100%), four corresponded to extraction replicates (33%), and two were not replicates and could correspond to contaminated samples (i.e., genetically similar individuals). However, for the 18 extraction replicates, twelve (66.66%) did not belong to the first distribution but were closer to their replicate than any other individual: 9/9 F1 (100% retrieved), 3/4 F2 (75%), and 0/5 backcrossed (because the five specimens clustered together).

Polarization of the alleles was performed. To do this, the reference allele (coded as 0) was chosen as the nucleotide with the highest proportion of *P. toxostoma* in wild populations. A particular case involves an SNP present only in a single homozygous state in *P. toxostoma* populations (coded as 0) and present in a single homozygous state (different) in *T. souffia* populations (coded as 2). In the following, such SNPs are referred to as diagnostic SNPs.

The definition of a diagnostic SNP was relaxed for discriminant SNPs by defining a score associated with each SNP as the mean average of the proportion of homozygous genotypes (0/0) in *P. toxostoma* individuals and the proportion of homozygous genotypes (1/1) in *T. souffia* individuals. This score varied between 0 and 1 and was exactly equal to 1 in the case of a diagnostic SNP. In the following, an SNP will be considered discriminant if its score is strictly higher than a fixed threshold (0.9). This threshold was chosen from the empirical score distribution (see Section 3).

From a theoretical point of view, an F1 genotype for a diagnostic SNP that presents a sum of alleles equal to 1 is heterozygous (0/1). To verify this on our empirical data, a volcanic eruption plot was generated considering discriminant SNPs. In this plot, a point refers to an SNP. The abscissa (x) shows the mean genotype score calculated on the F1-hybrid specimens, and the ordinate (y) shows the standard deviation of the genotype score calculated on the F1-hybrid specimens. The point of coordinates (x = 1, y = 0) indicates perfect agreement with the theory. A deviation from the value x = 1 was tested using the classical Student's t-test (type I error set at 0.05). All SNPs which deviated significantly from 1 were removed. When several SNPs had the same coordinate, the size of the point was proportional to the number of SNPs.

2.5 Principal component analysis

PCA was performed on the data corresponding to the selected 641 SNPs and 350 individuals. The original data corresponded to (Cij), $1 \leq i \leq n$, $1 \leq j \leq p$ where Cij was equal to 0, 1, or 2 and expressed the genotype of the *j*th SNP for an individual i. Here, *p* refers to the number of SNPs selected using previous procedures. When a value Cij was missing, it was

imputed by the mean genotypic value calculated on individuals belonging to the same group as the individual i. The normalization proposed by Patterson et al. (2006) was used before performing PCA. Moreover, the test of significance for axes was performed following the procedure of Patterson et al. (2006).

2.6 Bayesian procedure

We used a Bayesian model-based clustering algorithm implemented in the software STRUCTURE version 2.3.4 (Hubisz, Falush, Stephens, & Pritchard, 2009). Individuals in the sample were assigned to K populations or a mix of populations if their genotypes indicated that they were admixed (i.e., for F1, F2, and backcross). For the ancestry model, we used the option 'admixture model' because different alleles for hybrid specimens (F1, F2, and backcross) could be inherited from different species populations. Furthermore, we selected the option 'allele frequencies correlated'. The burn-in length was set to 100,000 followed by 1,000,000 iterations using a Markov Chain Monte Carlo as recommended previously (Pritchard et al., 2000). These analyses were carried out with K = 2 (i.e., two species).

2.7 Genomic combination of F2-hybrid specimens

Considering the 204 F2-hybrid specimens, we observed important mortality (38/204, 18.63%) occurring during 2 weeks, corresponding to a 2-cm size. Because this event could correspond to a specific post-zygotic barrier underlined by a particular genomic combination, we first analyzed the cohort of 166 F2 individuals who survived (>2 cm), and second, the cohort of 38 F2 individuals who died very young (=2 cm). For both cohorts, we tested two hypotheses related to the gametic and zygotic phases, respectively, during the formation of the F2 genome. Our study corresponded to a segregation distortion analysis (Fu et al., 2020).

The first hypothesis was related to allelic transmission from F1 to F2; if allele transmission is independent of allele origin, then H0_1 (hypothesis H0 number 1): the mean genotype value is equal to 1, for F2-hybrid specimens as for F1s. Otherwise, the genotype mean value is different from the value 1 (H1_1). This test was performed using the Student's t-test with a type I error set at 0.05. A rejection of H0_1 may be interpreted as a selection of alleles in favor of *P. toxostoma* or *T. souffia*. The second hypothesis was related to the allelic association to form genotypes in panmixia. The hypothesis H0_2 indicates that the genotypes are in Hardy–Weinberg (HW) equilibrium, while H1_2 indicates HW disequilibrium. We used the exact Hardy–Weinberg test (Weir, 1996), using the HWExactMat function of the HardyWeinberg package (Graffelman, 2015). Disequilibrium may correspond to several situations: an excess of heterozygous, an excess of homozygous, or an excess of homozygous loci linked to a species (*P. toxostoma* or *T. souffia*). We used the D chi-square statistic from the HWChisqMat function from the HardyWeinberg package (Graffelman, 2015) to identify the different situations. Statistical analyses were performed using R software version 3.6.3 (R Core Team, 2020).

2.8 Gene ontology (GO term) analysis on orthologous ENSDARGs

To perform overrepresentation of biological functions, we used the PANTHER classification system (http://www.pantherdb.org/) with a reference constituted by the 617 SNPs corresponding to 542 orthologous ENSDARGs. We further performed Fisher's exact test and false discovery rate correction. Because significant differences could be found between individuals belonging to the first cohort >2 cm and those belonging to the cohort = 2 cm, two separate analyses were performed.

2.9 Ethics statement

All sampling and experimental protocols involving the fish in this study (*Telestes souffia* and *Parachondrostoma toxostoma*) were reviewed and approved by local regulatory agencies: the AFB (Agence Française pour la Biodiversité) and the DDT (Direction Départementale des Territoires) from Alpes-de-Haute-Provence, Hautes-Alpes and Vaucluse (authorization numbers 2007–573 and 2008–636), following national regulations.

3 Results

3.1 Selection of discriminant loci using the *P. toxostoma* orthologous ENSDARGs

Based on the parameters described in the Material and Methods, we conserved all the individuals reported in Table 1. The initial data set included 369 individuals related to 15 groups and 5,289 SNPs. The repeatability of 34 specimen replicates (DNA and tissue) within a run and between runs was verified, and 18 replicates were removed (see Section 2). Figure 2 displays the missing (NA) individual rate by a group. Two clusters could be observed: the first one was constituted by samples yielding around 50% NA and the other one by samples yielding 10% NA. This was clearly related to the sequencing depth of the two runs. We removed one specimen that presented a very high level

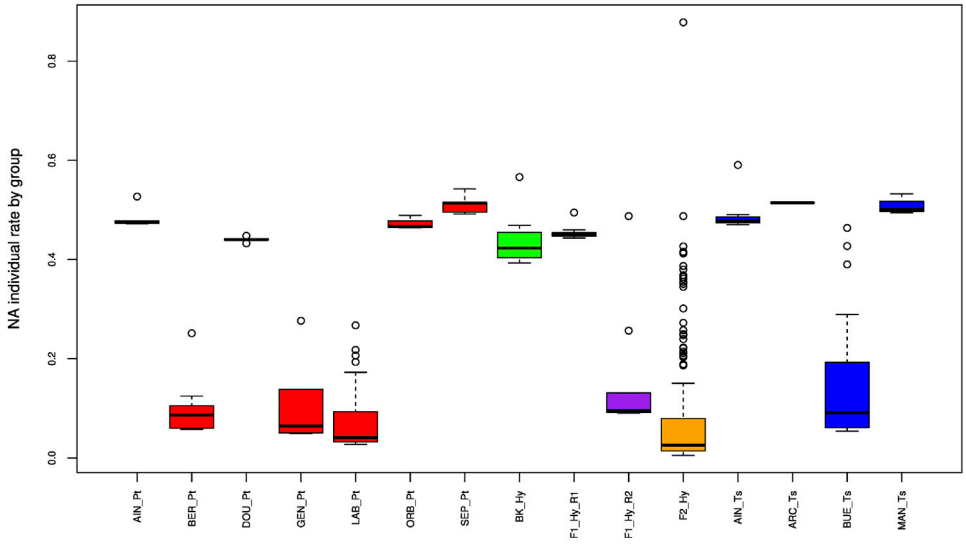


FIGURE 2
Boxplot of the missing data rate for the different groups (see Table 1). In red, *P. toxostoma* populations (AIN_Pt for Ain River, DOU_Pt for Doubs River, ORB_Pt for Orbieu River, BER_Pt for Berre River, GEN_Pt for genitors, LAB_Pt for laboratory specimens, SEP_Pt for Serre-Ponçon Lake); in blue, *T. souffia* populations (AIN_Ts for Ain River, ARC_Ts for Archidiacre site on Durance River, BUE_Ts for Buech River, MAN_Ts for Manosque on Durance River); in purple, F1_Hy_R1 for first run of F1 laboratory specimens, and F1_Hy_R2 for second run of F1 laboratory specimens; in orange, F2_Hy for F2 specimens; in green, BK_Hy for backcrossed laboratory specimens between F1 and *P. toxostoma* laboratory specimens.

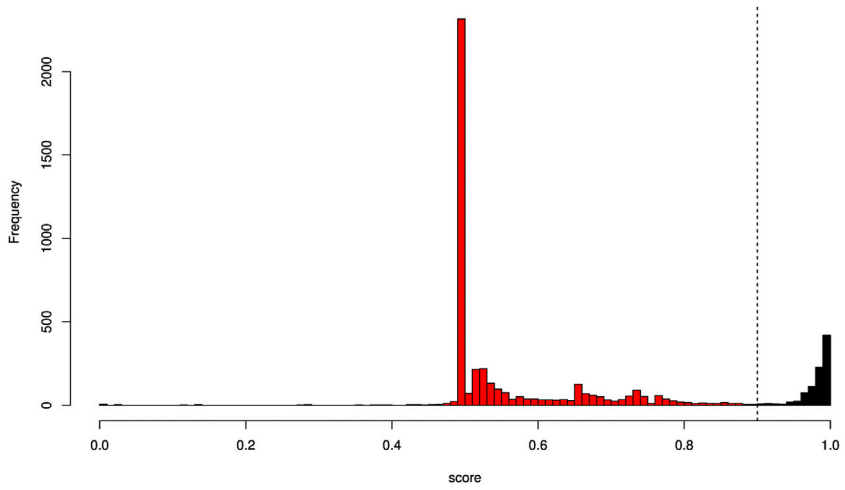


FIGURE 3
Histogram of SNP discriminating score. The *x*-axis represents the score values; the *y*-axis represents the absolute frequency of SNP. The vertical dotted line indicates the threshold (0.9) considered to define an SNP as discriminant. SNPs whose scores were higher than 0.9 were conserved and qualified as discriminant (in black). SNPs with scores less than or equal to 0.9 (in red) were discarded.

(>80%) of NA. This individual was observed in the deeper sequencing run (second run).

We then calculated the discriminant score for each SNP considering the two species *P. toxostoma* and *T. souffia*. The SNP discriminant score ranged from 0 to 1. An SNP exhibiting a score of 1 corresponds to a diagnostic locus. When the score is less than 1 and higher than 0.5, the loci is considered semi-diagnostic, and a score less than or equal to 0.5 refers to a non-diagnostic locus. We used a stringent cut-off of 0.9 (Figure 3) and restricted the dataset to the 908 SNPs considered

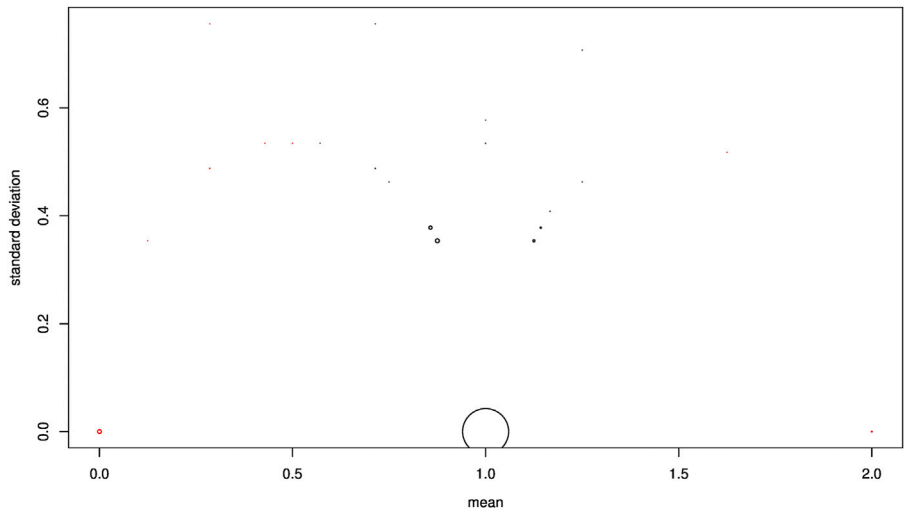


FIGURE 4
Volcano eruption plot. Each SNP is represented by a dot. The abscissa shows the mean genomic score calculated on F1 specimens; the ordinate shows the standard deviation of genomic scores calculated on F1 specimens. A coordinate point (x = 1, y = 0) is in perfect agreement with the Hardy–Weinberg equilibrium, indicating that an F1 individual is a heterozygote. Any point that deviates from (x = 1, y = 0) deviates from the HW equilibrium, indicating at least one homozygous F1 individual. All SNPs that deviate significantly from 1 on the *x*-axis are indicated in red; these SNPs were then discarded from the rest of the study. When several SNPs had the same coordinate, the size of the point was proportional to the number of SNPs.

discriminant loci. The discriminant SNPs represented 17% of the 5,289 SNPs. These SNPs were slightly related to chromosomes (Chi^2 = 42.14, df = 24, p = 0.011), with a higher rate of discriminant SNPs for chromosomes Chr5 and Chr20 (see Supplementary Figure S1 for more information). However, chromosome Chr3 had a lower rate of discriminant SNPs.

To validate the Mendelian inheritance of discriminant SNPs, we analyzed each 908 selected loci in F1-hybrids. As 204 SNPs were missing for the whole F1-hybrids, these were removed. The volcanic eruption plot (Figure 4) shows, for each SNP, the mean score value for the F1-hybrids, and the corresponding standard deviation on the ordinates. A majority of loci (507/704) presented the expected F1 values (i.e., heterozygous genotypes for all loci), with a mean score of 1 on the *x*-axis and a standard deviation of 0 on the *y*-axis. The mean score for 63 of the 704 loci deviated significantly from 1 (*t*-test, $p < 0.05$), and was removed. Finally, we obtained an intermediate data set constituted of 641 SNPs and 350 individuals. The rest of the SNP corresponded to shared alleles between the two species. The 641 SNPs corresponded to 560 *P. toxostoma* orthologous ENSDARGs, showing a decreasing exponential distribution. Indeed, 492 of the ENSDARGS were related to one SNP, 58 to two SNPs, 7 to three SNPs, and 3 to four SNPs. The genetic architecture map for pure individuals of each species, F1-hybrids, F2-hybrids, and backcross is presented in Figure 5.

3.2 Quantification of genomic dilution for the different hybrid specimens

PCA based on 350 individuals and 641 SNPs clearly displayed phylogenetic information on the first axis (54.24%, $p < 10^{-6}$) (Figure 6). Indeed, the specimens belonging to the pure species are defined by the first axis: groups belonging to *P. toxostoma* species DOU_Pt, SEP_Pt, AIN_Pt (corresponding to parapatric populations), BER_Pt, ORB_Pt (corresponding to allopatric populations), GEN_Pt (from the Durance River), and LAB_Pt (corresponding to *P. toxostoma* laboratory stock); and those belonging to *T. souffia* ARC_Ts, MAN_Ts, BUE_Ts, and AIN_Ts. The F1-hybrids displayed a central position with low dispersion, indicating a classical 50/50 genomic composition. Similarly, the F2-hybrids displayed a central position but showed a wide range between the two species, indicating genomic heterogeneity due to parental dilution. On the other hand, as expected, the backcross specimens occupied an intermediate position between the F1-hybrid and *P. toxostoma* specimens.

These results were in agreement with the STRUCTURE analysis with k = 2 (Supplementary Figure S2). The groups belonging to the *P. toxostoma* species were assigned to one cluster, and those belonging to *T. souffia* were assigned to another cluster. The Q-score of the F1-hybrid specimens for the *P. toxostoma* and *T. souffia* clusters presented a mean of

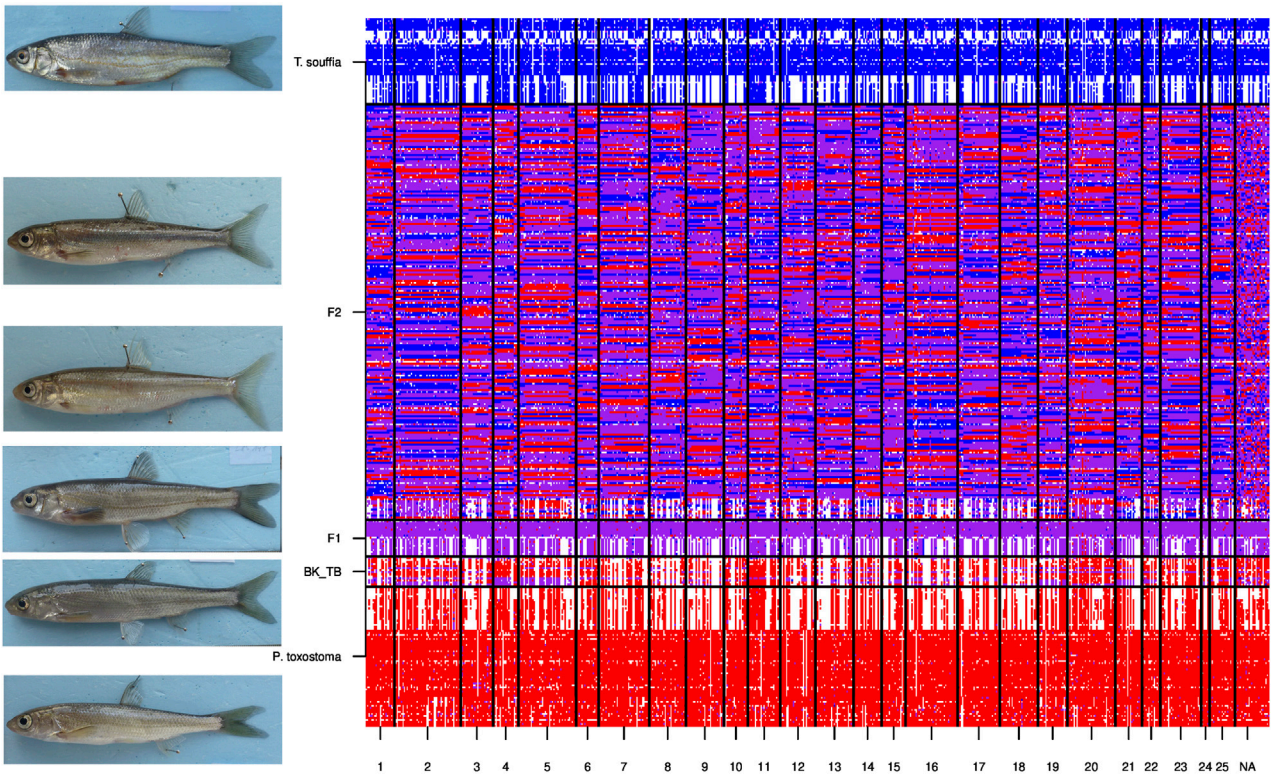


FIGURE 5
Genetic architecture of nuclear DNA SNPs for hybrid and parental populations. Lines represent specimens belonging to the six groups: respectively from bottom to top: *P. toxostoma*, Backcrosses (*P. Toxostoma* X F1-hybrids), F1-hybrids, F2-hybrids and *T. souffia*. Columns correspond to the 641 discriminants SNPs following chromosoms numbers (1 to 25, NA for not attributed). Each intersection line X column corresponds to a genotype. This genotype is indicated in red for homozygous *P. toxostoma*, in blue for homozygous *T. souffia*, in purple for heterozygous, and in white for missing genotype.

0.507 and 0.493, respectively, and a standard error of 0.008, indicating a similar genotype. The F2-hybrid displayed a mean of 0.513 (0.487) and a standard error of 0.06, indicating a mosaic of genotypes. Finally, the backcrossed specimens yielded a mean of 0.866 (0.134) and a standard error of 0.122, indicating a genomic introgression to *P. toxostoma*.

3.3 Establishment of endogenous selection in F2 specimens

We next considered 204 F2-hybrids specimens and the 617 SNPs among 641 associated with *Danio rerio* chromosomes. However, to explore a potential postzygotic barrier (lineage sorting of particular genomic combination), we first analyzed the cohort of 166 F2 individuals who survived (>2 cm) and, second, the cohort of 38 F2 individuals who died very young (=2 cm). Of the 617 discriminant SNPs, only 86 SNPs (13.94%) were significantly ($p < 0.05$) different from allele frequencies $p = q = 0.5$, indicating an increase of allele frequencies toward one species (Figure 7A). Interestingly, some genomic regions were clearly weighted toward *P. toxostoma,* especially in the chromosomes Chr5 and Chr20, while those weighted toward *T. souffia* were more sparse. The allele frequency ranged from 0.55 to 0.84 for *P. toxostoma* and 0.55 to 0.65 for *T. souffia*.

In the second test, we tested the HW equilibrium for each locus. We found that a majority of loci (563/617, 91.25%) were in HW equilibrium. The loci departing from HW equilibrium mainly presented an excess of heterozygotes (47/617, 7.62%) (Supplementary Figure S3A).

Figure 8A shows the relationship between these two phenomena, that is, the deviation of the allele frequencies and HW disequilibrium. A large majority (475/617 loci, 76.99%) displayed a neutral evolution, with HW equilibrium and no deviation in allele frequency, indicating high genomic compatibility between the two genomes. We could consider the absence of deviation proportion as an estimate of the porosity between the two species. Furthermore, 45 loci (7.29%) presented

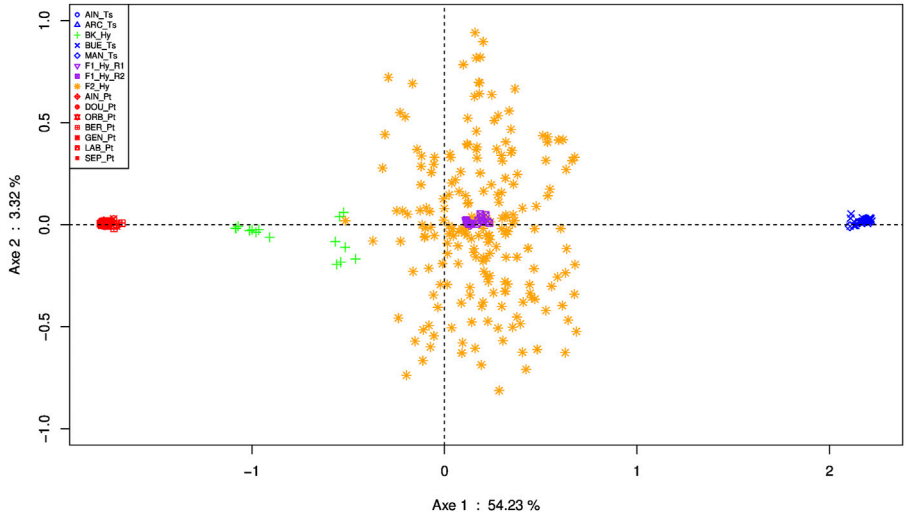


FIGURE 6
Specimen representation on the main PCA plane performed on SNP discriminant data, with standardization as proposed by (Patterson et al., 2006). In red, *P. toxostoma* populations (AIN_Pt for Ain River, DOU_Pt for Doubs River, ORB_Pt for Orbieu River, BER_Pt for Berre River, GEN_Pt for genitors, LAB_Pt for laboratory specimens, SEP_Pt for Serre-Ponçon Lake); in blue, *T. souffia* populations (AIN_Ts for the Ain River, ARC_Ts for the Archidiacre site in Durance River, BUE_Ts for the Buech River, MAN_Ts for Manosque on the Durance River); in purple, F1_Hy_R1 for the first series of F1 laboratory samples, and F1_Hy_R2 for the second series of F1 laboratory samples; in orange, F2_Hy for F2 samples; in green, BK_Hy for laboratory samples backcrossed between F1 and *P. toxostoma*.

no deviation in allele frequency between the two parental species but were in HW disequilibrium. 42 loci (6.81%) presented an excess of heterozygous genotypes (purple triangle, Figure 8A), indicating a heterozygous advantage. These loci were mainly related to three chromosomes: Chr3, Chr9, and Chr15 (Supplementary Figure S3A). In contrast, only three loci (0.49%) presented a heterozygous disadvantage (green square, Figure 8A). Among the 84 loci (13.61%) presenting a higher frequency toward *P. toxostoma* (red color in Figure 8A), 76 were in HW equilibrium (red circle in Figure 8A), 4 presented excess homozygous *P. toxostoma* genotypes (red square in Figure 8A), and 4 presented an excess of heterozygous genotypes (red triangle in Figure 8A). Among the 13 loci (2.11%) presenting a higher frequency toward *T. souffia* (blue color in Figure 8A), 12 displayed HW equilibrium (blue circle in Figure 8A), and 1 presented an excess of heterozygous genotypes (blue triangle in Figure 8A).

Next, we focused our analysis on the cohort of F2s who died very young (=2 cm). This analysis displayed a very different pattern, as shown in Figure 7B. Of the 617 discriminant SNP, only 59 SNP (9.56%) were significantly different from allele frequencies $p = q = 0.5$. We detected genomic regions weighted toward *P. toxostoma* (chromosomes Chr16 and Chr21) and spotted regions (Chromosomes 7, 10, and 20), combined with genomic regions weighted toward *T. souffia* (Chromosome 11) and a spotted chromosome (Chr15). This result clearly indicated a selection toward *T. souffia* alleles in Chromosomes 11 and 15. However, the HW disequilibrium was less pronounced for this cohort than in the cohort >2 cm (Figure 8B, Supplementary Figure S3B), indicating that most loci were in HW equilibrium. Indeed, 9 loci (purple triangle) presented an excess of heterozygous genotypes, and 2 (green square) presented a heterozygous disadvantage.

The porosity between the two genomes appears to be high but not complete. Indeed, we observed the signature of endogenous selection in favor of both species, mostly toward *P. toxostoma*. This pattern was particularly marked on the cohort > 2 cm indicating the presence of filters limiting the genome porosity during fish development. Considering F2 genotypes, we also observed a positive selection in favor of heterozygotes corresponding to an increase in the genomic porosity by maintaining high genomic compatibility.

3.4 Gene function associated with the two F2 cohorts

Of the 142 genes that presented an atypical pattern, that is, deviation of allele frequencies and/or HW disequilibrium, in the cohort >2 cm, two categories were well represented. The 76 genes that presented allelic frequency toward *P. toxostoma* and HW equilibrium (red circle in Figure 8) were implicated in positive regulation of developmental processes (GO:0,051,094; 5/7; p = 5.08E-03), regulation of organelle organization (GO:

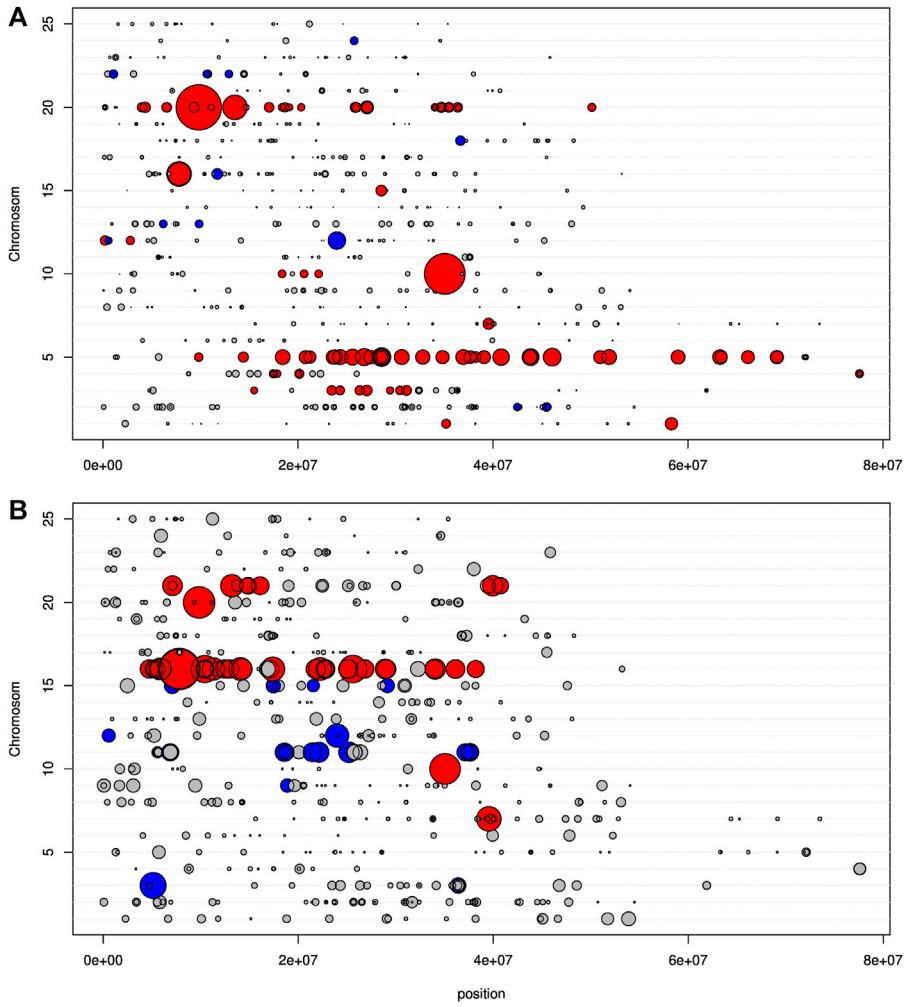


FIGURE 7
Representation on *D. rerio* chromosomes of the t-statistic expressing the deviation from 1 for the mean genotypic score, calculated on F2 individuals. In grey, non-significant deviation; in blue, significant deviation toward *T. souffia*; in red, significant deviation toward *P. toxostoma*. The size of the dot is proportional to the deviation. **(A)** Specimens with a size >2 cm. **(B)** Specimens with a size of 2 cm.

0,033,043; 6/17; p = 2.67E-02, sptbn2 and cntrl), especially in the cytoskeleton (GO:0,051,493; 5/14; p = 4.14E-02), chromosome organization (GO:0,051,276; 5/14; p = 4.14E-02, mcm3l and recql), cell cycle (GO:0,007,049; 5/14; p = 4.14E-02, ptpa), and negative regulation of transcription, DNA-templated (GO:0,045,892; 5/15; p = 5.08E-02, hif1ab). The 42 genes that presented an excess of heterozygous genotypes (purple triangle in Figure 8) were implicated in several developmental functions, such as innate immune response (GO:0,045,087; 2/4; p = 7.07E-02, stat1b, ifih1), pronephros development (GO:0,048,793; 2/4; 7.05E-02, prkcsh), and liver development (GO:0,001,889; 2/5; 9.40E-02, leg1a).

Of the 70 genes that presented an atypical pattern in the cohort = 2 cm, two categories were well represented. Indeed, 35 corresponded to allelic frequency toward *P. toxostoma* and HW equilibrium (red circle in Figure 8), and 15 corresponded to allelic frequency toward *T. souffia* and HW equilibrium (blue circle in Figure 8). The first cluster was implicated in endosome to lysosome transport (GO:0,008,333; 2/2; p = 3.76E-03, trak1a), mitochondrion organization (GO:0,007,005; 2/2; p = 3.76E-03, ulk1b), vesicle cytoskeletal trafficking (GO:0,099,518; 2/5; p = 1.34E-02, myo1eb), and regulation of protein depolymerization (GO:1,901,879; 2/5; p = 1.71E-02, nestin). The second cluster was implicated in positive regulation of potassium ion transport (GO:0,043,268; p = 1.40E-02, wnk3) and positive regulation of sodium ion transport (GO:0,010,765; p = 1.40E-02, wnk2).

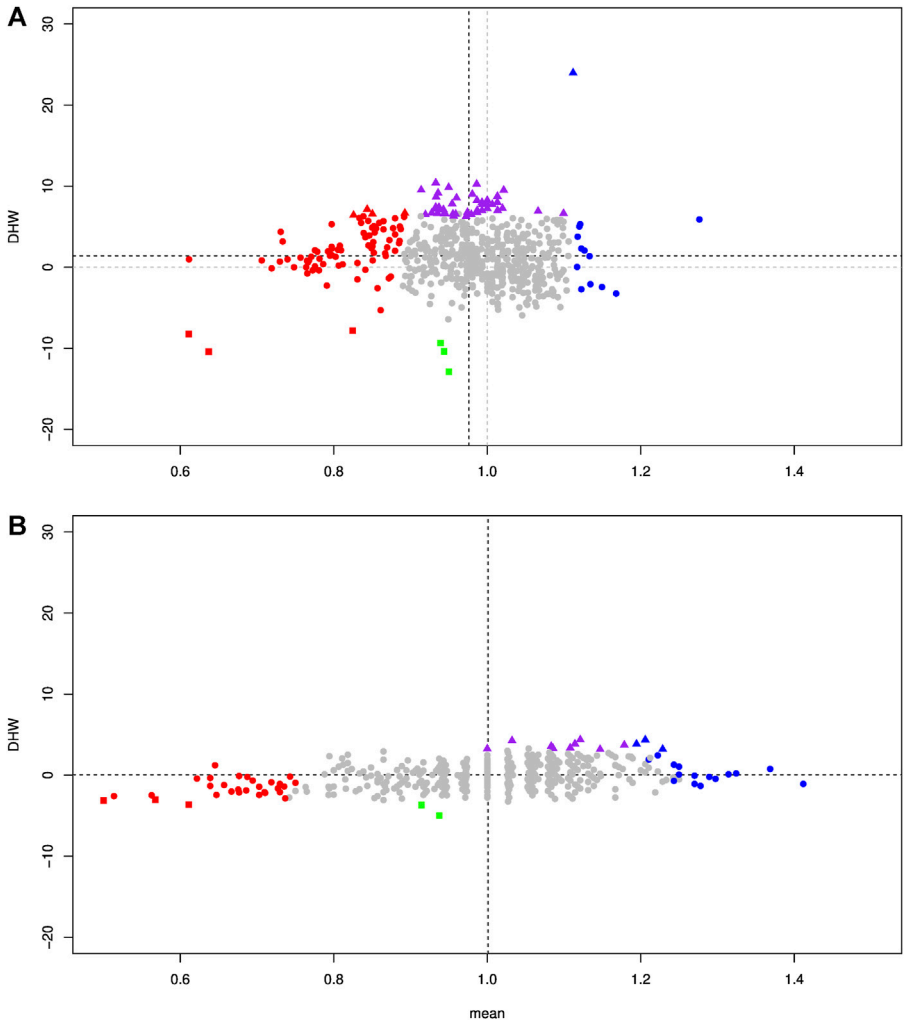


FIGURE 8
Biplot of the DHW statistic of the Hardy–Weinberg deviation as a function of the t-statistic expressing the deviation from 1 for the allelic composition, calculated on F2 individuals. Grey circles represent the SNPs that were not significant for both tests. Red circles and blue circles represent SNPs in HW equilibrium, for which allele composition deviates significantly from 1 toward *P. toxostoma* and *T. souffia*, respectively. Red and blue triangles represent SNPs in Hardy–Weinberg disequilibrium (excess heterozygote) for which allelic composition deviates significantly from 1 toward *P. toxostoma* and *T. souffia*, respectively. Purple triangles represent SNPs in Hardy–Weinberg disequilibrium (excess heterozygote) whose allele composition does not deviate significantly from 1. Red squares represent SNPs in Hardy–Weinberg disequilibrium (excess *P. toxostoma* homozygote) whose allelic composition deviates significantly from 1 toward *P. toxostoma*. Green squares represent SNPs in Hardy–Weinberg disequilibrium (excess homozygote for both species) for which allelic composition does not deviate significantly from 1. **(A)** Specimens with a size >2 cm. **(B)** Specimens with a size of 2 cm.

4 Discussion

4.1 The importance of studying new hybrid models in laboratory: reproductive isolation versus genetic incompatibility

The importance of performing crosses in the laboratory to study speciation/hybridization is no longer in question (Liu et al., 2016; White et al., 2020). However, some biological models, such as many European river cyprinids, require several years of rearing before breeding, especially F2-hybrid specimens, and typically survival rate data is based on F1-hybrid specimens (Philippart and Berrebi, 1990; Nzau Matondo et al., 2007). At first sight, we expected a very strong genetic incompatibility between these two species because the reproductive isolation seemed particularly important given the anecdotal number of hybrids (n = 2) observed in the field for more than 25 years (Dubut et al., 2010; Dubut et al., 2012). However, we have observed a strong and repeatable genomic compatibility in our farms. It is, therefore, crucial to describe

and quantify this compatibility in order to explain the decoupling between the reproductive isolation observed in the wild and the strong genetic compatibility in the laboratory of two such divergent species living in sympatry.

4.2 F1 individuals are the cornerstone in the calibration of discriminant single-nucleotide polymorphisms

The first strength of our work was that we were able to select discriminant SNPs from different parental populations. This allowed us to take into account the polymorphism of each species by having an *a priori* knowledge of the existence of hybrid individuals, as they have been the subject of numerous studies (Dubut et al., 2010; Dubut et al., 2012; Corse et al., 2015). The second strength of our work was the use of F1 individuals, which allowed us to select the discriminant SNPs of the parental populations that were indeed in the heterozygous state in the F1s. We found that a large majority of these SNPs (507/704, 72.02%) were consistently heterozygous in F1s. Among them, each of the 415 SNP was associated with a single transcript (conveniently named orthologous ENSDARG), and their distribution across the *Danio* transcriptome/genome was not completely random ($Chi^2 = 42.14$, df = 24, $p = 0.011$). Two explanations can be formulated: The first is that our *P. toxostoma* transcriptome has non-random areas for which we did not obtain the genes, which results in missing some SNPs. The second is that by using the position of the *D. rerio* genes, we assumed that the position of the genes on the chromosomes studied had not changed. It is clear that large regions of synteny exist in Cyprinidae (Liu et al., 2017a; Chen et al., 2019; Lei et al., 2021); however, there are also shuffles (Liu et al., 2017b; Chen et al., 2019). If these shuffles are not random, then they may explain a preferential association between SNPs, ENSDARGs, and chromosomes.

4.3 Remodeling a hybrid genome: Allelic frequency and Hardy–Weinberg disequilibrium in F2s

The selected SNPs show a very strong discriminating power. Whether using PCA or assignment by STRUCTURE, we clearly separated the different genetic groups: *P. toxostoma*, *T. souffia*, F1-hybrids, F2-hybrids, and backcross. We were thus able to focus on the F2-hybrid individuals that survived beyond 2 cm (i.e., >2 cm). The study of the 617 SNPs showed that a large majority (475/617 loci, 76.99%) were evolving under a panmictic model; that is, the alleles of these genes were interchangeable. The set of F2 individuals that we analyzed constitutes a sorting of compatibility between our two species after a succession of pre- and post-zygotic stages, such as gamete recognition, fertilization, embryonic development, and larval metamorphosis. Two phenomena are underlined by our results.

The first phenomenon is related to the significant deviation of the allelic frequency ($p \neq q$, directional selection) in the F2 specimens, notably in favor of *P. toxostoma*, although it was also present in favor of *T. souffia*. The fact that the extreme majority of these loci were in HW equilibrium implies an unbalanced production of gametes, whose association would be random in aquatic environments. Moreover, we note that certain genomic regions, such as chromosomes Chr5 and Chr20, seemed to display this phenomenon. If these regions are indeed regions of synteny between *D. rerio* and the two other Cyprinidae species (*P. toxostoma* and *T. souffia*), then the structural incompatibility may involve a problem of chromatin interaction in these regions. This phenomenon could be linked with the asymmetric expression of alleles in cyprinid hybrid lineages (Ren et al., 2016; Ren et al., 2019). Furthermore, in 2-cm-sized individuals with high mortality, we were able to observe particular genomic signatures displaying this phenomenon but involving different genomic regions (chromosomes Chr16 and Chr21).

The second is related to HW disequilibrium, essentially genes with frequencies $p = q = 0.5$. In this case, we believe that this phenomenon is the result of successive sorting during the different stages of development, which will continue until adulthood. Some genomic combinations will be favored (heterozygous advantage and stabilizing selection) while others will be selected against (disruptive selection).

4.4 What are the functions of these genes?

We were able to identify two broad categories of genes. Genes that tended to be more *P. toxostoma*-like, while being in HW equilibrium, were involved in developmental regulatory processes and cytoskeletal organization but were also involved in chromosome organization and DNA regulation, which suggests not only that epigenetic factors are of great importance in the development and maintenance of the hybrids but also that this stability is more related to *P. toxostoma* alleles. Genes that showed a heterozygous advantage were involved in immune response and kidney/liver development.

Individuals with a 2-cm size, which died almost at the same time, had common characteristics, notably in endosome to lysosome transport and cytoskeletal trafficking, with allele frequency toward *P. toxostoma*. However, these F2-hybrids presented differences in genes involved in mitochondrion organization and positive regulation of potassium and sodium. In conclusion, the 'genomic landscape' of F2-hybrids is a mixture of genes whose alleles are fully

compatible, at least under laboratory conditions, and genes whose alleles show different evolutionary trajectories (Martin and Jiggins, 2017).

4.5 Potential shortcomings and limitations

Our study showed that two endemic species belonging to two different genera had very high genomic porosity with some disadvantaged genomic combinations. We were able to identify certain genes and biological functions associated with these disadvantages. Currently, it is not possible to know whether these patterns result from a structural problem related to the DNA sequences due to allele differences (*T. souffia* versus *P. toxostoma*) or the regulation of these genes. In the first case, it would be interesting to use the complete set of SNP markers to describe the different selections between gene and intergenic regions. However, the use of the *Danio rerio* genome remains limited because of the low similarity between *Telestes/Parachondrostoma* sequences and the *Danio rerio* genome, and a reference *Telestes/Parachondrostoma* genome would be necessary. In the second case, it would be necessary to work on RNA sequencing to analyze differential gene expression (DGE) in order to estimate the effects of dominance, co-dominance, and recessivity of the different alleles belonging to *T. souffia* and *P. toxostoma*.

Moreover, it will be interesting to test the parameters involved in the overlapping breeding because the reciprocal crosses (i.e., one female *P. toxostoma* and one male *T. souffia*) did not produce hybrid individuals. Currently, we don't know if this result is due to a genetic incompatibility or a handling problem because nine backcrosses were obtained by crossing female *P. toxostoma* and male F1-hybrids, with *P. toxostoma* mitochondrial DNA (i.e., *T. souffia* for F2-hybrids). These results may be the consequence of imprinting effects.

4.6 Can we expect the emergence of a *P. toxostoma*–*T. souffia* hybrid zone?

The results obtained in the laboratory show that the main barrier is not related to genomic incompatibility. We, therefore, deduce that the barriers which currently strongly limit the emergence of a *P. toxostoma*–*T. souffia* hybrid zone are exogenous. To establish the emergence of a *P. toxostoma*–*T. souffia* hybrid zone, the pre-zygotic barrier would have to be broken between the two species (Crispo et al., 2011; Grabenstein and Taylor, 2018), forcing them to reproduce on the same spawning grounds at the same time, as observed for other cyprinids (Balon, 1992). *P. toxostoma* specimens reproduce essentially during the month of April until the beginning of May in the range of water temperature from 11 to 13°C in the Buech tributary of the Durance, France (Costedoat et al., 2007), whereas *T. souffia* specimens reproduce in June, in water temperatures from 10 to 14°C. Thus, an increase in river temperature in late spring, coupled with a decrease in river flow, could be the trigger for natural hybridization between these two species, as in a faster warming environment, the breeding periods would overlap. It can then be assumed that the embryonic development of the hybrids would exhibit *T. souffia*-like embryonic development. Indeed, the embryonic development is 7 days at 13.9°C for *T. souffia*, while the embryonic development of *P. toxostoma* is 8 days at 16°C and 10 days at 14°C (Gozlan et al., 1999). However, in adulthood, the advantage would be in favor of *P. toxostoma*, because the optimal adult temperature range for *P. toxostoma* is 16–25°C, while it is between 10 and 18°C for *T. souffia* (Tissot and Souchon, 2010). There would thus be adaptive trade-offs related to growth, reproduction, feeding, and river temperature (Daufresne et al., 2004), which could be implemented by the selection of new hybrid combinations (exogenous selection). Obviously, these trends would be evolutionary compromises depending on the genes presenting genetically compatible alleles (endogenous selection) and the density of each species. It would not be surprising to observe an asymmetrical introgression in such cyprinid fish hybrid zones (Crespin et al., 1999). With climate change, future studies should focus on the establishment of new hybrid zones involving a genomic mixing of several endemic and/or introduced species (common garden or *in natura*). It will be particularly interesting to decipher the gene expression of F1 and F2 hybrid specimens considering their genomic combination, especially for the F2 specimens (evolutionary novelties, transgressive segregation, etc.). Common garden experiments will be associated with field sampling in order to detect new hybrid specimens that underwent the filter of environmental selection. They will have to better understand the part related to endogenous selection from those related to exogenous selection. These topics are essential in the context of global warming and accelerated environmental changes.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=863842, accession number PRJNA863842.

Ethics statement

All sampling and experimental protocols involving the fish in this study (Telestes souffia and Parachondrostoma toxostoma) were reviewed and approved by local regulatory agencies: AFB (Agence Française pour la Biodiversité) and the DDT (Direction

Départementale des Territoires) from Alpes-de-Haute-Provence, Hautes-Alpes and Vaucluse, authorization number 2007–573 and 2008–636, following national regulations.

Author contributions

AG and NP developed the approach and contributed to the conception of the study. AG, FD, NP, and YT analyzed data. AG and NP wrote the manuscript. AG and BB developed mesocosms. AG and J-FM developed molecular experiments. RC managed fieldwork.

Funding

This work was supported by the UMR 1467 RECOVER *Aix* Marseille Univ. INRAE.

Acknowledgments

The authors would like to thank Sylvain Santoni and Audrey Weber, UMR AGAP Montpellier FRANCE, for molecular biology support. They thank Azélie Buisson very much for generating the species distribution map. Many thanks go to the three referees for their very constructive comments. They would like to thank Editage (www.editage.com) for English language editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.963341/full#supplementary-material

References

Anderson, E. (1948). Hybridization of the habitat. *Evolution* 2, 1–9. doi:10.1111/j.1558-5646.1948.tb02726.x

Balon, E. K. (1992). How dams on the River Danube might have caused hybridization and influenced the appearance of a new cyprinid taxon. *Environ. Biol. Fishes* 33, 167–180. doi:10.1007/bf00002562

Brazier, T., Cherif, E., Martin, J. F., Gilles, A., Blanchet, S., Zhao, Y., et al. (2021). A tale of an invader: Reconstructing the genomic history of invasive top mouth gudgeon (*Pseudorasbora parva*) populations. Available at: https://www.authorea.com/users/397838/articles/510606-a-tale-of-an-invader-reconstructing-the-genomic-history-of-invasive-topmouth-gudgeon-pseudorasbora-parva-populations?commit=0c5e5849916cb7227bc278dcb584a8ba31e40997 (Accessed February 20226.

Canestrelli, D., Porretta, D., Lowe, W. H., Bisconti, R., Carere, C., and Nascetti, G. (2016). The tangled evolutionary legacies of range expansion and hybridization. *Trends Ecol. Evol.* 31, 677–688. doi:10.1016/j.tree.2016.06.010

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi:10.1111/mec.12354

Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., et al. (2019). De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Sci. Adv.* 5, eaav0547. doi:10.1126/sciadv.aav0547

Chunco, A. J. (2014). Hybridization in a warmer world. *Ecol. Evol.* 4, 2019–2031. doi:10.1002/ece3.1052

Corse, E., Pech, N., Sinama, M., Costedoat, C., Chappaz, R., and Gilles, A. (2015). When anthropogenic river disturbance decreases hybridisation between non-native and endemic cyprinids and drives an ecomorphological displacement towards juvenile state in both species. *PLOS ONE* 10, e0142592. doi:10.1371/journal.pone.0142592

Costedoat, C., Pech, N., Chappaz, R., and Gilles, A. (2007). Novelties in hybrid zones: Crossroads between population genomic and ecological approaches. *PLOS ONE* 2, e357. doi:10.1371/journal.pone.0000357

Crespin, L., Berrebi, P., and Lebreton, J-D. (1999). Asymmetrical introgression in a freshwater fish hybrid zone as revealed by a morphological index of hybridization. *Biol. J. Linn. Soc. Lond.* 67, 57–72. doi:10.1111/j.1095-8312.1999.tb01929.x

Crispo, E., Moore, J. S., Lee-Yaw, J. A., Gray, S. M., and Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays* 33, 508–518. doi:10.1002/bies.201000154

Daufresne, M., Roger, M. C., Capra, H., and Lamouroux, N. (2004). Long-term changes within the invertebrate and fish communities of the Upper Rhône River: Effects of climatic factors. *Glob. Chang. Biol.* 10, 124–140. doi:10.1046/j.1529-8817.2003.00720.x

Díaz-Arce, N., and Rodríguez-Ezpeleta, N. (2019). Selecting RAD-seq data analysis parameters for population genetics: The more the better? *Front. Genet.* 10, 533. doi:10.3389/fgene.2019.00533

Doadrio, I., and Carmona, J. A. (2004). Phylogenetic relationships and biogeography of the genus Chondrostoma inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 33, 802–815. doi:10.1016/j.ympev.2004.07.008

Dowling, T. E., and DeMarais, B. D. (1993). Evolutionary significance of introgressive hybridization in cyprini fishes. *Nature* 362, 444–446. doi:10.1038/362444a0

Dubut, V., Fouquet, A., Voisin, A., Costedoat, C., Chappaz, R., and Gilles, A. (2012). From late miocene to holocene: Processes of differentiation within the Telestes genus (actinopterygii: Cyprinidae). *PLOS ONE* 7, e34423. doi:10.1371/journal.pone.0034423

Dubut, V., Sinama, M., Martin, J. F., Meglécz, E., Fernandez, J., Chappaz, R., et al. (2010). Cross-species amplification of 41 microsatellites in European cyprinids: A tool for evolutionary, population genetics and hybridization studies. *BMC Res. Notes* 3, 135. doi:10.1186/1756-0500-3-135

Edmands, S. (2002). Does parental divergence predict reproductive compatibility? *Trends Ecol. Evol.* 17, 520–527. doi:10.1016/S0169-5347(02)02585-5

Fu, Q., Meng, X., Luan, S., Chen, B., Cao, J., Li, X., et al. (2020). Segregation distortion: High genetic load suggested by a Chinese shrimp family under high-intensity selection. *Sci. Rep.* 10, 21820. doi:10.1038/s41598-020-78389-w

Gozlan, R. E., Copp, G. H., and Tourenq, J. N. (1999). Early development of the sofie, *Chondrostoma toxostoma*. *Environ. Biol. Fishes* 56, 67–77. doi:10.1023/a:1007576222907

Grabenstein, K. C., and Taylor, S. A. (2018). Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol. Evol.* 33, 198–212. doi:10.1016/j.tree.2017.12.008

Graffelman, J. (2015). Exploring diallelic genetic markers: The Hardy-Weinberg package. *J. Stat. Softw.* 64, 1–23. doi:10.18637/jss.v064.i03

Hayden, B., McLoone, P., Coyne, J., and Caffrey, J. M. (2014). Extensive hybridisation between roach, Rutilus Rutilus L., and common bream, Abramis brama L., in Irish lakes and rivers. *Biol. Environ.* 114B, 35–39. doi:10.3318/bioe.2014.04

Jansson, G., Thulin, C. G., and Pehrson, Å. (2007). Factors related to the occurrence of hybrids between Brown hares Lepus europaeus and mountain hares *L. timidus* in Sweden. *Ecography* 30, 709–715. doi:10.1111/j.2007.0906-7590.05162.x

Konopiński, M. K., and Amirowicz, A. (2018). Genetic composition of a population of natural common bream *Abramis brama* × roach *Rutilus rutilus* hybrids and their morphological characteristics in comparison with parent species. *J. Fish. Biol.* 92, 365–385. doi:10.1111/jfb.13506

Kottelat, M. (1997). European freshwater fishes. An heuristic checklist of the freshwater fishes of Europe (exclusive of former USSR), with an introduction for non-systematists and comments on nomenclature and conservation. *Biol. Bratisl.* 52, 1–271.

Lei, Y., Yang, L., Jiang, H., Chen, J., Sun, N., Lv, W., et al. (2021). Recent genome duplications facilitate the phenotypic diversity of Hb repertoire in the Cyprinidae. *Sci. China. Life Sci.* 64, 1149–1164. doi:10.1007/s11427-020-1809-0

Lewontin, R. C., and Birch, L. C. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution* 20, 315–336. doi:10.1111/j.1558-5646.1966.tb03369.x

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. Arxiv: Genomics, 1–3. doi:10.48550/arXiv.1303.3997

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liu, H., Chen, C., Gao, Z., Min, J., Gu, Y., Jian, J., et al. (2017a). The draft genome of blunt snout bream (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet. *GigaScience* 6, 1–13. doi:10.1093/gigascience/gix039

Liu, H., Fu, B., Pang, M., Feng, X., Yu, X., and Tong, J. (2017b). A high-density genetic linkage map and QTL fine mapping for body weight in crucian carp (*Carassius auratus*) using 2b-RAD sequencing. *G3 (Bethesda)* 7, 2473–2487. doi:10.1534/g3.117.041376

Liu, S., Luo, J., Chai, J., Ren, L., Zhou, Y., Huang, F., et al. (2016). Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1327–1332. doi:10.1073/pnas.1512955113

Martin, S. H., and Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* 47, 69–74. doi:10.1016/j.gde.2017.08.007

Matute, D. R., Butler, I. A., Turissini, D. A., and Coyne, J. A. (2010). A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329, 1518–1521. doi:10.1126/science.1193440

Moyle, L. C., and Nakazato, T. (2010). Hybrid incompatibility "snowballs" between Solanum species. *Science* 329, 1521–1523. doi:10.1126/science.1193063

Nzau Matondo, B., Ovidio, M., Poncin, P., Kakesa, T. A., Wamuini, L. S., and Philippart, J. C. (2007). Hybridization success of three common European cyprinid species, *Rutilus rutilus*, *Blicca bjoerkna* and *Abramis brama* and larval resistance to stress tests. *Fish. Sci.* 73, 1137–1146. doi:10.1111/j.1444-2906.2007.01445.x

Orr, H. A., and Turelli, M. (2001). The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55, 1085–1094. doi:10.1111/j.0014-3820.2001.tb00628.x

Owens, G. L., and Samuk, K. (2020). Adaptive introgression during environmental change can weaken reproductive isolation. *Nat. Clim. Chang.* 10, 58–62. doi:10.1038/s41558-019-0628-0

Oziolor, E. M., Reid, N. M., Yair, S., Lee, K. M., Guberman VerPloeg, S., Bruns, P. C., et al. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* 364, 455–457. doi:10.1126/science.aav4155

Paris, J. R., Stevens, J. R., and Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods Ecol. Evol.* 8, 1360–1373. doi:10.1111/2041-210X.12775

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190. doi:10.1371/journal.pgen.0020190

Perea, S., Böhme, M., Zupančič, P., Freyhof, J., Šanda, R., Özuluğ, M., et al. (2010). Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evol. Biol.* 10, 265. doi:10.1186/1471-2148-10-265

Pereira, C. S. A., Aboim, M. A., Ráb, P., and Collares-Pereira, M. J. (2014). Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity* 112, 343–350. doi:10.1038/hdy.2013.110

Philippart, J. C., and Berrebi, P. (1990). Experimental hybridization of *Barbus barbus* and Barbus meridionalis: Physiological, morphological, and genetic aspects. *Aquat. Living Resour.* 3, 325–332. doi:10.1051/alr:1990033

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Ren, L., Li, W., Qin, Q., Dai, H., Han, F., Xiao, J., et al. (2019). The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* × Culter Alburnus. *Genome Res.* 29, 1805–1815. doi:10.1101/gr.249805.119

Ren, L., Li, W., Tao, M., Qin, Q., Luo, J., Chai, J., et al. (2016). Homoeologue expression insights into the basis of growth heterosis at the intersection of ploidy and hybridity in Cyprinidae. *Sci. Rep.* 6, 27040. doi:10.1038/srep27040

Rhymer, J. M., and Simberloff, D. (1996). Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* 27, 83–109. doi:10.1146/annurev.ecolsys.27.1.83

Rochette, N. C., and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12, 2640–2659. doi:10.1038/nprot.2017.123

Roule, L. (1925). *Les Poissons des eaux douces de la France*. Paris: Les Presses Universitaires Fr, 340.

Runemark, A., Vallejo-Marin, M., and Meier, J. I. (2019). Eukaryote hybrid genomes. *PLOS Genet.* 15, e1008404. doi:10.1371/journal.pgen.1008404Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6880984/

Schumer, M., Cui, R., Rosenthal, G. G., and Andolfatto, P. (2015). Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet.* 11, e1005041. doi:10.1371/journal.pgen.1005041

Scribner, K. T., Page, K. S., and Bartron, M. L. (2000). Hybridization in freshwater fishes: A review of case studies and cytonuclear methods of biological inference. *Rev. Fish. Biol. Fish.* 10, 293–323. doi:10.1023/A:1016642723238

Smith, G. R. (1992). Introgression in fishes: Significance for paleontology, cladistics, and evolutionary rates. *Syst. Biol.* 41, 41–47. doi:10.2307/2992505

Stelkens, R., and Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63, 884–897. doi:10.1111/j.1558-5646.2008.00599.x

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., and Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. *Curr. Biol.* 24, 671–676. doi:10.1016/j.cub.2014.01.069

Tissot, L., and Souchon, Y. (2010). Synthèse des tolérances thermiques des principales espèces de Poissons des rivières et fleuves de plaine de l'ouest européen. *Hydroécol. Appl.* 17, 17–76. doi:10.1051/hydro/2010004

Ungaro, A., Pech, N., Martin, J. F., McCairns, R. J. S., Mévy, J. P., Chappaz, R., et al. (2017). Challenges and advances for transcriptome assembly in non-model species. *PLOS ONE* 12, e0185020. doi:10.1371/journal.pone.0185020

Weir, B. S. (1996). "Genetic data analysis II," in *Methods for discrete population genetic data* (Massachusetts, United States: Sinauer), 466.

White, N. J., Snook, R. R., and Eyres, I. (2020). The past and future of experimental speciation. *Trends Ecol. Evol.* 35, 10–21. doi:10.1016/j.tree.2019.08.009

You, M., and Xu, J. (2021). What are the best parents for hybrid progeny? An investigation into the human pathogenic fungus cryptococcus. *J. Fungi* 7, 299. doi:10.3390/jof7040299

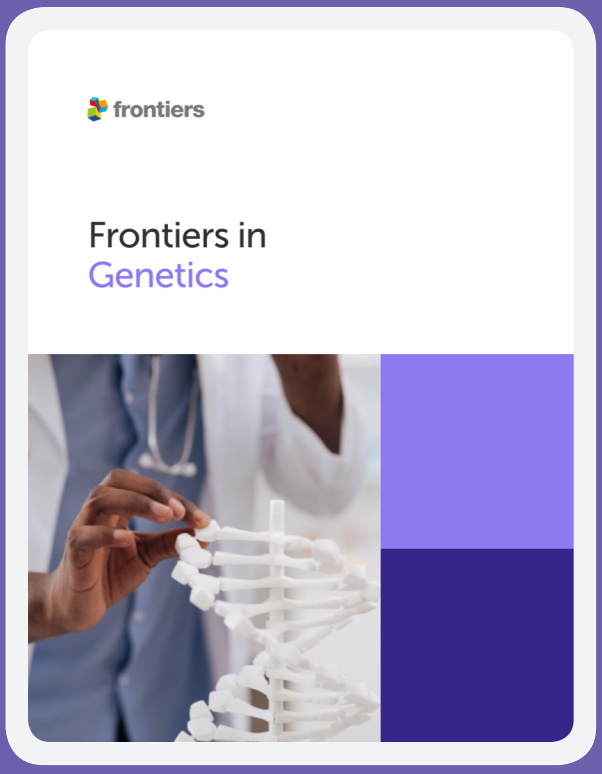
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact