




Article

Genetic Diversity and Population Differentiation of Kashgarian Loach (*Triplophysa yarkandensis*) in Xinjiang Tarim River Basin

Xiaoyun Zhou ¹, Shaokui Yi ^{2,*}, Wenhao Zhao ¹, Qiong Zhou ¹, Jianzhong Shen ¹, Dapeng Li ¹, Bin Huo ¹ and Rong Tang ¹

¹ College of Fisheries, Key Laboratory of Freshwater Animal Breeding, Ministry of Agriculture and Rural Affairs, Huazhong Agricultural University, Wuhan 430070, China; zhoxuy@mail.hzau.edu.cn (X.Z.); 2019308110039@webmail.hzau.edu.cn (W.Z.); hainan@mail.hzau.edu.cn (Q.Z.); jzhsh@mail.hzau.edu.cn (J.S.); ldp@mail.hzau.edu.cn (D.L.); huobin@mail.hzau.edu.cn (B.H.); tangrong@mail.hzau.edu.cn (R.T.)
² College of Life Sciences, Huzhou University, Huzhou 313000, China
* Correspondence: 02844@zjhu.edu.cn

Simple Summary: The distribution of Kashgarian loach (*Triplophysa yarkandensis*) is limited to the Tarim River basin, which is the largest inland river in China. However, the population size of *T. yarkandensis* has been diminishing, and it is critically endangered in the Tarim River basin due to the gradual depletion of water resources, together with alien invasion and agricultural cultivation in Tarim River. In this study, we adopted the RAD-seq method to investigate the population genetics of *T. yarkandensis*, and a high degree of genetic variations and significant genetic differentiation was detected among *T. yarkandensis* populations in the Tarim River basin. The obtained data contribute to understanding the genetic status of *T. yarkandensis*, and help to provide the scientific management strategies and direct future monitoring and utilization of the genetic resource in Xinjiang region.



Citation: Zhou, X.; Yi, S.; Zhao, W.; Zhou, Q.; Shen, J.; Li, D.; Huo, B.; Tang, R. Genetic Diversity and Population Differentiation of Kashgarian Loach (*Triplophysa yarkandensis*) in Xinjiang Tarim River Basin. *Biology* **2021**, *10*, 734. <https://doi.org/10.3390/biology10080734>

Academic Editor: De-Li Shi

Received: 15 June 2021

Accepted: 27 July 2021

Published: 1 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The distribution of *Triplophysa yarkandensis* is restricted to Xinjiang's Tarim River basin. We collected 119 *T. yarkandensis* samples from nine geographic populations in the Tarim River basin and utilized the RAD-seq method for SNP genotyping. In this study, a total of 164.81 Gb bases were generated with the Illumina platform, and 129,873 candidate SNPs were obtained with the Stacks pipeline for population genetic analyses. High levels of genetic diversity were detected among nine populations. The AMOVA results showed that the majority of genetic variations originated from among populations ($F_{ST} = 0.67$), and the pairwise F_{ST} values ranged from 0.4579 to 0.8736, indicating high levels of genetic differentiation among these populations. The discriminate analysis of principal components (DAPCs) and neighbor joining (NJ) tree revealed that the nine populations could be separated into two clusters (i.e., south and north populations), and modest genetic differentiation between south and north populations was observed, while the individuals from several populations were not clustered together by geographical location. The evidence of two genetic boundaries between south and north populations (except TTM) was supported by barrier analysis. The Bayesian skyline plotting indicated that *T. yarkandensis* populations in the Tarim River basin had not experienced genetic bottlenecks, and the effective population size remained stable. This study first clarified the genetic diversity and differentiation of *T. yarkandensis* populations in the Tarim River basin, and it provided valuable molecular data for conservation and management of natural populations.

Keywords: Tarim River; gene flow; barrier; Bayesian skyline plotting

1. Introduction

Triplophysa yarkandensis, belonging to Nemacheilidae, is a kind of small economic fish. It is mainly distributed in slow-moving and shallow water habitats. Specifically, the distribution of *T. yarkandensis* is limited to the Tarim River basin, which is the largest inland river in China. The Tarim River and its tributaries are typical seasonal rivers in an arid area,

and it often does not produce runoff. In recent years, climate change and human activities resulted in the water curtailment and river desiccation of the Tarim River basin [1]. As a result, due to the gradual depletion of water resources, together with alien invasion and agricultural cultivation in the Tarim River, the population size of *T. yarkandensis* has been diminishing and it is critically endangered [2]. Changing this trajectory will require coordinated research and conservation strategies to provide a better understanding of population genetics at large spatial scales.

T. yarkandensis plays an important role in the fragile local ecosystem of the Tarim River basin, and previous studies have shown that the origin and evolution of the genus of *Triplophysa* may be associated with uplifts of the Tibetan Plateau [3–5]. Previously, many studies have reported the distribution, morphological [6] and reproductive characteristics [7], mitochondrial genome [8,9] and transcriptome [2] of *T. yarkandensis*. However, information on population structure and historical population dynamics of *T. yarkandensis* populations in the Tarim River basin is still lacking. Therefore, evaluating the genetic diversity, population structure and population dynamics of *T. yarkandensis* is quite essential for the further utilization of genetic germplasm resources and wild population conservation.

Over the past years, restriction site-associated DNA sequencing (RAD-seq) has become a powerful and widely-used technique in conservation biology studies, especially for the species without a reference genome [10–12]. Unlike microsatellites and mitochondrial genes (e.g., *cytb*, *coi*, d-loop), more than thousands of SNPs at genomic level could substantially improve the capacity to unravel genetic diversity and divergence in species with both strong and weak population structure [13]. For instance, the significant genetic differentiation of *Misgurnus anguillicaudatus* populations in the Yangtze River basin was detected with 2092 high-quality SNPs [14], but another study failed to detect these findings using microsatellite and mitochondrial genes [15]. Recently, the RAD-seq technique, including GBS, 2b-RAD and dd-RAD, has been widely used in genetic studies of aquatic species, such as red swamp crayfish [16], black bass (*Micropterus spp.*) [17], *Rhodeus ocellatus kurumeus* [18] and Nuijiang catfish (*Creteuchiloglanis macropterus*) [11]. In this study, we used the RAD-seq method to unravel the population dynamics and population structure of *T. yarkandensis* in the Tarim River basin. We aimed to (1) use SNP genotyping to investigate fine-scale genetic variations of *T. yarkandensis* populations; (2) evaluate the population differentiation of the populations in the Tarim River basin and (3) provide genomic evidence on the historical population dynamics of *T. yarkandensis*. These data are essential to providing scientific management strategies and direct future monitoring of the genetic resource of *T. yarkandensis* in Xinjiang region.

2. Materials and Methods

2.1. Sampling and DNA Extraction

Fish samples were taken from 14 locations in the Tarim River basin using nylon line fishing nets from May to October 2019. In total, 119 specimens were collected from nine sample sites in the Tarim River basin (Figure 1). The tail fin clips from the specimens were obtained and stored in 95% ethyl alcohol. Total genomic DNA was extracted from 119 samples using the QIAamp[®] DNA Mini Kit (QIAGEN, Hilden, Germany) following manufacturer's instructions. The RNA pollution was treated using Ribonuclease A (TaKaRa, Dalian, China). DNA quality was assessed by 1.2% agarose gel electrophoresis, and the DNA concentration was determined by the NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and Qubit 2.0 (Life Technologies, Gaithersburg, MD, USA).

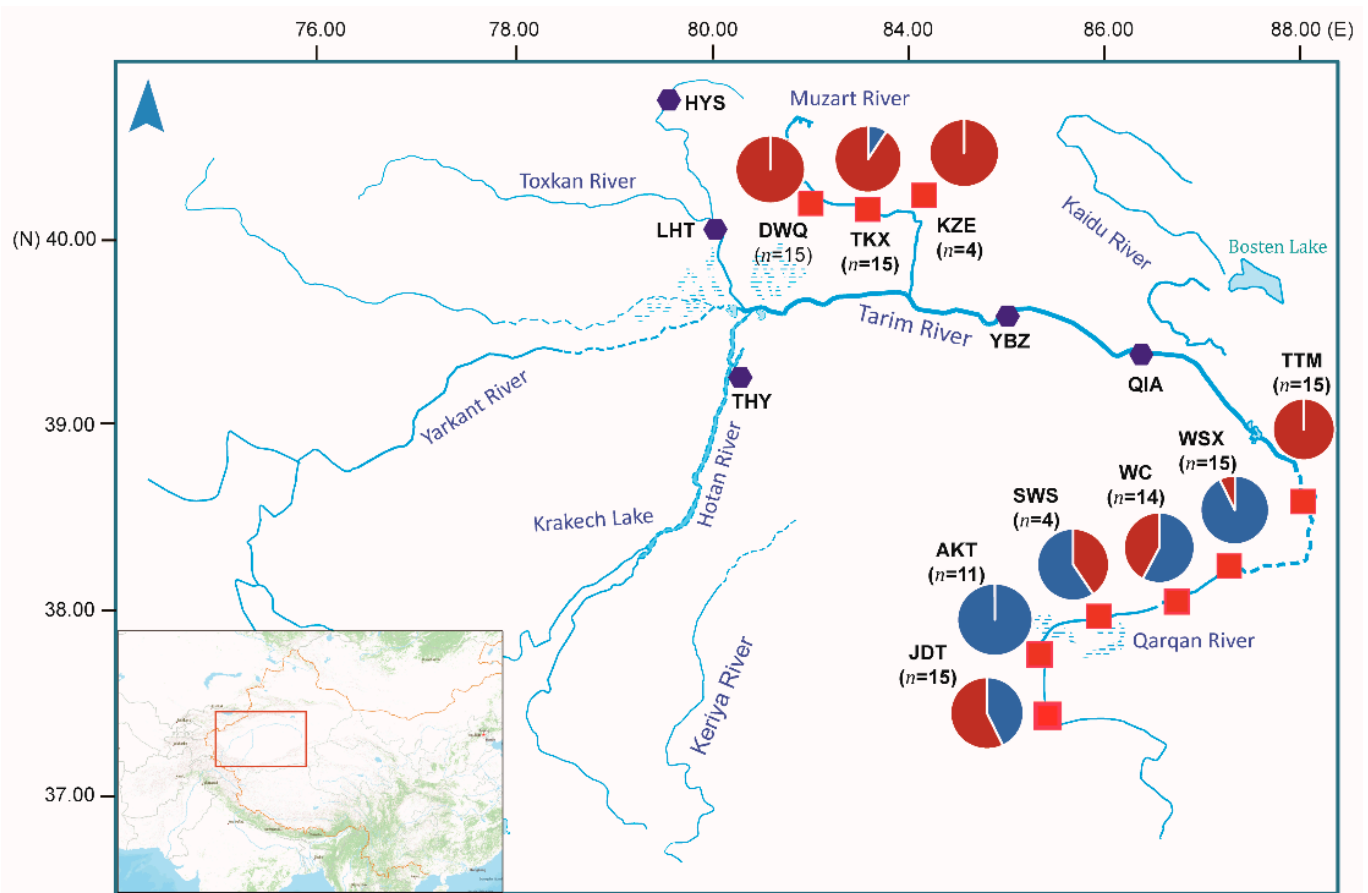


Figure 1. Sample sites of *T. yarkandensis* in the Tarim River basin located in Xinjiang. The hexagon symbol colored in dark purple represents the sampling locations where no specimen was collected. The red square symbols indicate the sampling locations used in this study. The pie in the map represents average admixture proportion of two genetically distinct clusters estimated with fastStructure.

2.2. RAD Library Construction and Sequencing

The predicted electronic enzymatic digestion was performed using the genome sequences of a closely-related species, *T. tibetana* [19], as a reference genome, and a combination of two endonucleases (*RsaI* and *HinCII*) was selected. Genomic DNA was incubated at 37 °C with *RsaI*, T4 DNA ligase, ATP, and *RsaI* adapter. Restriction-ligation reactions were heat-inactivated at 65 °C, and then digested for additional restriction enzyme *HinCII* at 37 °C. The PCR reaction was performed using diluted restriction-ligation samples, dNTP, Taq DNA polymerase and primer containing barcode. Subsequently, the PCR productions were purified and pooled. The pooled sample was incubated at 37 °C with *RsaI*, T4 DNA ligase, ATP and Solexa adapter. The digested fragment sequences with lengths of 364–464 bp were enriched for library construction following the protocol [20]. The PCR products were purified using a Gel Extraction Kit (QIAGEN, Hilden, German), and then the gel-purified products were sequenced on an Illumina HiSeq 2500 system (Illumina Inc., San Diego, CA, USA) with 125 paired-end.

2.3. Data Processing

The raw data generated from the Illumina sequencer were filtered with the fastp program [21]. To confirm the reliability of data, we randomly selected 2000 reads and mapped them to the NCBI nucleotide non-redundant nucleotide sequence (NT) database. The Stacks pipeline [22] was used for RAD-seq data processing, SNP discovery and standard population genetic statistics. Given that the reference genome of *T. yarkandensis* is not available, a de novo orthology search was conducted. Briefly, the retained reads were demultiplexed

according to the barcodes using the `process_radtags` program from the Stacks pipeline. The `ustacks` program was used to align the individuals into exactly matching stacks with the parameter of $M = 4$. A catalog of RAD loci was built with `cstacks` and all individuals in the population were matched against the catalog using the `sstacks` program. The `populations` program was used to filter loci ($r = 0.75$, $\text{min-maf} = 0.05$, $\text{max_obs_het} = 1$). To address linkage disequilibrium (LD) between SNPs and to account for potential ascertainment bias, only one SNP per RAD locus was randomly selected with the `populations` parameter of `–write_single_snp`. The test of Hardy–Weinberg equilibrium (HWE) for each population was performed with `populations` program. To eliminate the bias of loci out of HWE, a Perl script `filter_hwe_by_pop.pl` implemented in dDocent program [23] was used to filter the loci with the parameters of `–h 0.001, –c 0.25`.

2.4. Genetic Diversity and Population Structure

The obtained loci were used to evaluate population genetic diversity, including major allele frequency, observed heterozygosity (H_O) and expected heterozygosity (H_E), and nucleotide diversity. Pairwise F_{ST} among the nine populations were determined with Arlequin v3.4.5 [24]. Genetic clustering and admixture analyses were performed with the “LEA” package [25] in R. The best-fit number of clusters (K) was assessed by calculating an entropy criterion with cross-validation using the `snmf` function. Lower cross-entropy values indicate a better model fit. Meanwhile, population admixture was also evaluated with a Bayesian approach using `fastStructure` [26]. The `ChooseK.py` script was used to determine the optimal K value. Meanwhile, we analyzed the genetic structure using the non-model-based multivariate approach DAPC (discriminant analysis of principal components) implemented in the “`adegen`” package [27]. The optimal number of clusters was determined as the one for which the Bayesian information criterion (BIC) showed the lowest value.

We used `plink v 1.90` [28] to perform principal components analysis (PCA), and the scatter plotting was performed with the first and second components using “`ggplot2`” package [29] in R. We constructed a neighbor joining (NJ) tree to cluster the populations using the “`ape`” package [30] in R. In addition, the distribution of genetic variations was determined by analyses of molecular variance (AMOVA) using the Arlequin v3.4.5. The genetic barriers associated with each geographical location and population were investigated using Monmonier’s maximum-difference algorithm in BARRIER version 2.2 [31]. The pairwise genetic distances among nine populations were determined by $F_{ST}/(1-F_{ST})$.

3. Results

3.1. SNP Identification and Genotyping

A total of 169.92 Gb raw data was generated by sequencing of the RAD libraries, with an average of 1.37 Gb bases per individual. After quality control, 164.81 Gb clean data was remained for the subsequent analysis (Table S1). All of the randomly selected reads mapped to the sequences of fish species (e.g., *Cyprinus carpio*, *Danio rerio*) in the NCBI NT database, which indicated that the obtained reads were reliable. An average of 10.90 million tags were obtained for each individual. After filtering the stacks with low coverage, 186,838 assembled loci were genotyped with an average coverage per individual of 17.10. According to the filtering criterion that loci presented in >75% individuals in the populations and $\text{MAF} > 0.05$, 129,873 candidate SNP markers were obtained for the subsequent genetic analysis (Table 1). With the filter using the `populations` program, the mean length of the remaining loci was 284.15 bp. Transversions (Tv) outweighed transition polymorphisms (Ti), accounting for 53.58% of the SNP sites, with an observed Ti:Tv ratio of 0.87.

Table 1. The summary of loci identified with the Stacks pipeline among nine populations.

Population ID	Num of Samples	Samples Per Locus	All Sites	Variant Sites	Polymorphic Sites	Fixed Alleles
TKX	15	12.091	6,825,098	12,019	10,346	32
DWQ	15	13.407	25,761,023	58,354	50,117	1040
SWS	15	13.570	31,271,934	69,585	5955	739
WC	14	12.552	31,093,652	71,003	66,518	1033
JDT	15	13.624	33,647,835	77,904	73,819	1406
AKT	11	10.054	33,397,250	74,608	63,564	839
KZE	4	3.552	40,984,502	97,754	66,521	502
WSX	15	12.424	5,779,906	10,335	8984	9
TTM	15	13.532	25,175,340	57,971	53,848	673
Total	119		54,958,861	129,873		

Note: Fixed alleles indicate the private alleles only detected in a population.

3.2. Genetic Diversity and Population Structure

The genetic diversity and allele polymorphism of nine populations were generally consistent. Observed heterozygosity (H_O) across all populations ranged from 0.2260 to 0.2604 within nine populations. H_O was lowest in WSX (0.2260) and highest in JDT (0.2604). The nucleotide diversity (P_i) among nine populations ranged from 0.2364 to 0.2861, which was highest in KZE. The individual inbreeding coefficients (F_{IS}) of the populations ranged from 0.0236 to 0.0779 (Table 2). The number of loci that varied from 0 to 4056 showed significant deviations from the Hardy–Weinberg equilibrium among nine populations. To investigate the genetic structure of the nine populations, DAPC was performed with inferred genetic clusters ($K = 1 \sim 9$). The DAPC BIC showed the lowest value with $K = 2$ (Figure 2A), and indicated that the optimal clustering solution was two genetic clusters (Figure 2B). Likewise, population admixture analysis was performed with the “snmf” function in the “LEA” package, and the number of clusters (K) was set from 1 to 9. The results revealed that the optimal number of clusters was two ($K = 2$) as identified by minimum cross-entropy value. In addition, the loci filtered by HWE test were used for the Bayesian analysis using fastStructure ($K = 1 \sim 9$). The result revealed that the optimal K value was 2 by running the ChooseK.py script, which is consistent with the number of clusters determined with “snmf” and DAPC methods. The average admixture proportions of two clusters for each population are shown in Figure 1. Most individuals from TKX, DWQ, KZE and TTM exhibited obviously different ancestry information compared with the individuals in the AKT, SWS, JDT, WSX and WC (Figure 2C). Several individuals from JDT and SWS also had the admixture ancestry component that dominated in TKX, DWQ and KZE.

Table 2. The genetic diversity and allele polymorphism of nine populations.

Pop ID	p	Obs_Het	Obs_Hom	Exp_Het	Exp_Hom	P_i	F_{IS}	HWE
TKX	0.8372	0.2356	0.7644	0.2345	0.7655	0.2446	0.0292	372
DWQ	0.8227	0.2462	0.7538	0.2508	0.7492	0.2607	0.0441	3299
SWS	0.8118	0.2562	0.7438	0.2686	0.7315	0.2789	0.0645	3606
WC	0.8107	0.2526	0.7474	0.2689	0.7311	0.2802	0.0779	3766
JDT	0.8078	0.2604	0.7396	0.2725	0.7275	0.2830	0.0667	4056
AKT	0.8244	0.2554	0.7446	0.2492	0.7508	0.2625	0.0236	1989
KZE	0.8195	0.2488	0.7512	0.2445	0.7555	0.2861	0.0715	0
WSX	0.8445	0.2260	0.7740	0.2269	0.7731	0.2364	0.0403	447
TTM	0.8139	0.2516	0.7484	0.2649	0.7351	0.2751	0.0714	3235

Note: p indicates mean frequency of the most frequent allele at each locus in this population. Obs_Het and Exp_Het represent mean observed and expected heterozygosity in this population, respectively. P_i indicates mean value of π in this population. HWE indicates the number of loci found to be significantly out of Hardy–Weinberg equilibrium ($p < 0.05$).

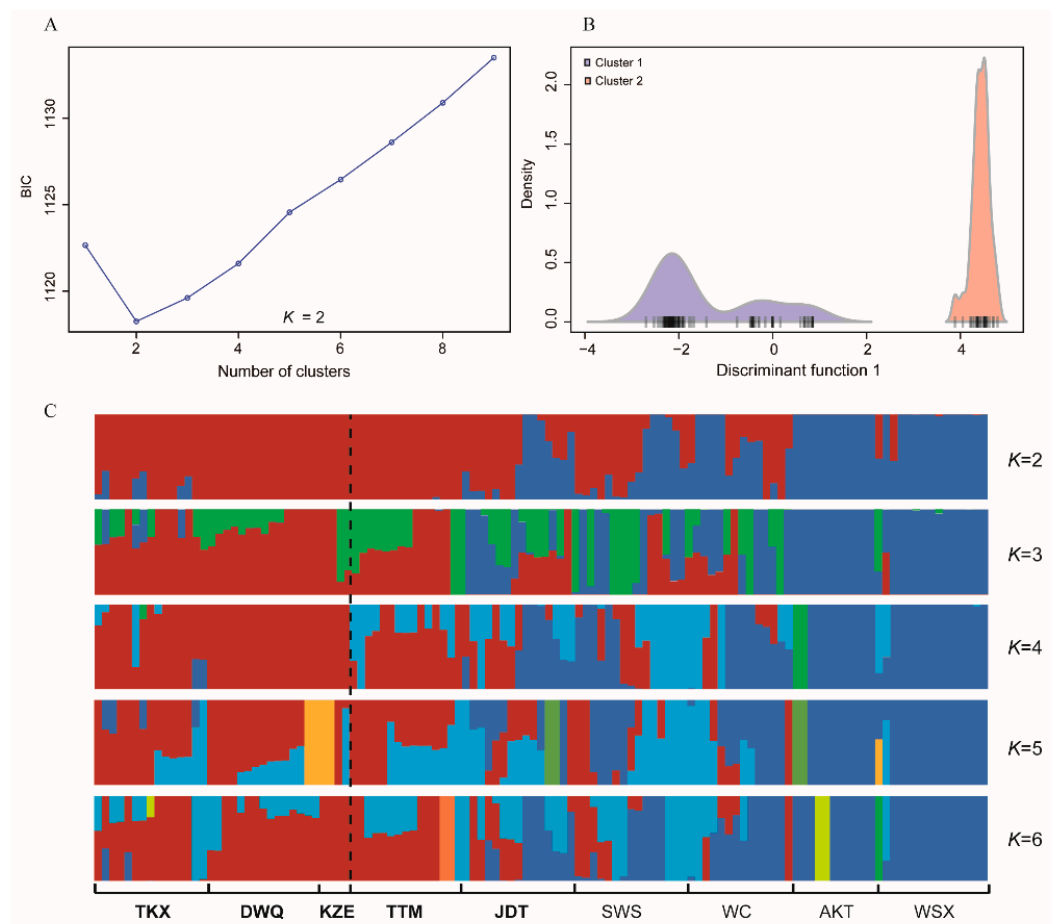


Figure 2. The distribution of BIC values with number of clusters ranged from 1 to 9 (A) and the densities of individuals on a given discriminant function for two clusters (B); the admixture clustering results for $K = 2\sim 6$ with Bayesian analysis using fastStructure based on the filtered dataset (C). The dashed line represents the separation of south and north populations.

The NJ tree of the 119 individuals (Figure 3A) features two genetically distinct clusters, with the most individuals from AKT, WC, WSX, and SWS grouping together (cluster 1) and the other individuals clustering into another cluster (cluster 2). The individuals from DWQ together with TKX and KZE individuals were dominant in the cluster 2, which coincided with the results of the structure analysis. Notably, the TTM individuals and some JDT individuals were clustered into cluster 2, and some TTM and JDT individuals were grouped into a subcluster in cluster 2, indicating a relatively high genetic distance between these two populations and other populations in cluster 1. Additionally, similar results were observed with the results of PCA plotting (Figure 3B). The PCA showed no significant clustering by populations or geographic locations, but the individuals from TKX, DWQ and KZE were grouped together, which have a smaller geographic distance than other populations. The individuals from AKT, WC and WSX together with some individuals from TTM, JDT and SWS were clustered together, and other individuals from TTM, JDT and SWS showed closer affinity to TKX, DWQ and KZE.

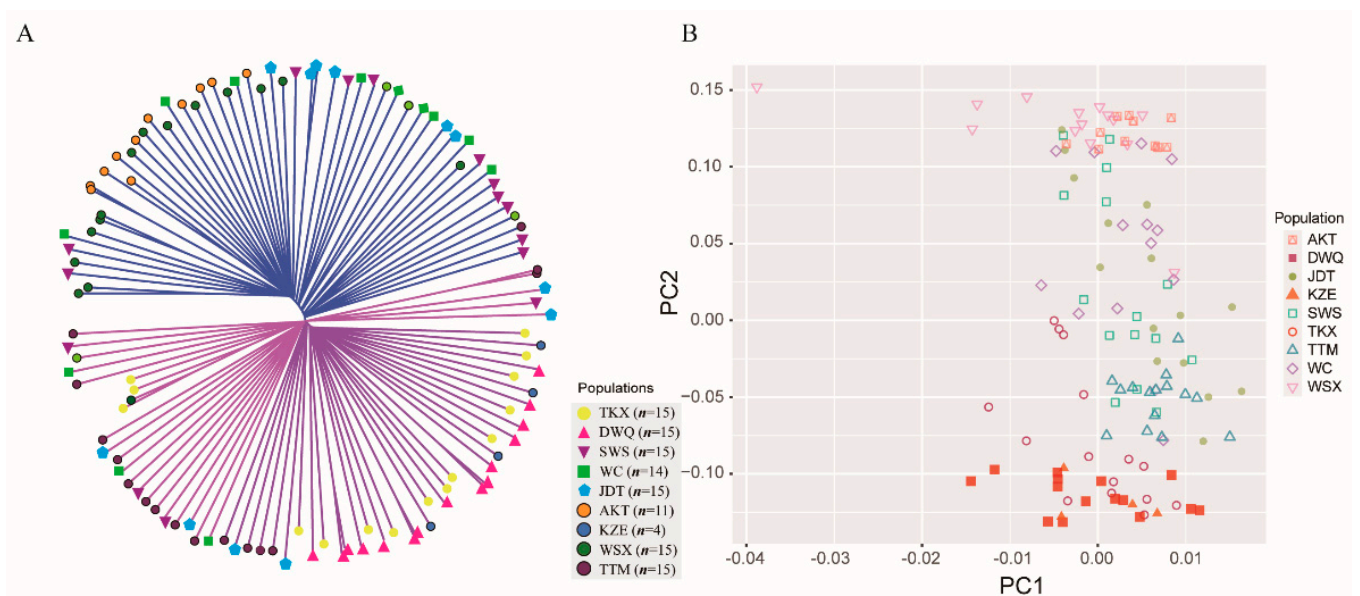


Figure 3. The NJ tree of 119 individuals constructed with the ape package (A) and the PCA plotting of the individuals with first and second principal components (B).

3.3. Gene Flow and Genetic Differentiation

Pairwise F_{ST} values across nine populations ranged from 0.4579 to 0.8736 (Table 3), indicating that the genetic differentiation among these populations was at very high levels. To trace the genetic variations among the populations, AMOVA was performed (Table S2). The result showed that the majority of genetic variation was due to between-population differentiation (67.01%) with evidence of significant genetic differentiation ($p < 0.05$). The association between genetic and geographic distance was tested with Barrier software, with improved Monmonier's method. The result showed that the genetic boundaries between the populations supported the two boundaries (Figure 4). The first boundary was between population TTM and south populations (WC, JDT, SWS, AKT and WSX), and the second boundary was found between north populations (DQW, TKX and KZE) and those south populations.

Table 3. Pairwise F_{ST} among nine populations of *T. yarkandensis*.

F_{ST}	TKX	DWQ	SWS	WC	JDT	AKT	KZE	WSX	TTM
TKX	0								
DWQ	0.7514	0							
SWS	0.7728	0.5961	0.0000						
WC	0.7720	0.6102	0.4823	0.0000					
JDT	0.7831	0.6126	0.4811	0.4579	0.0000				
AKT	0.8110	0.6397	0.5278	0.5318	0.4965	0.0000			
KZE	0.8617	0.6616	0.5944	0.5630	0.5294	0.5987	0.0000		
WSX	0.6625	0.7732	0.7826	0.7799	0.7903	0.8186	0.8736	0.0000	
TTM	0.7656	0.6207	0.5622	0.5957	0.5773	0.5976	0.6604	0.7719	0

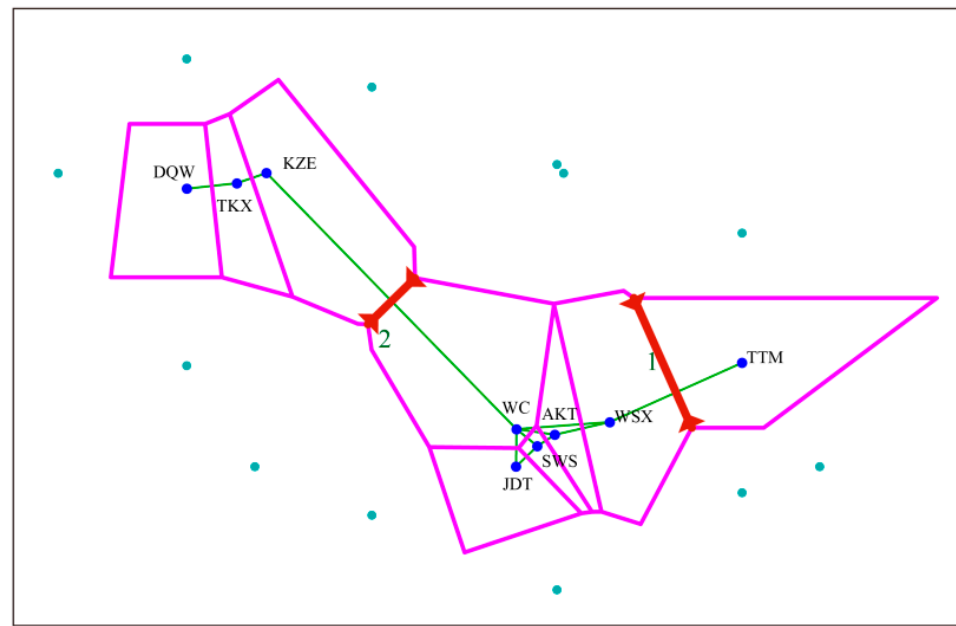


Figure 4. The two genetic boundaries (red bold lines) detected by BARRIER version 2.2 using genetic distance (N_m). Pink lines and solid green lines represent the Voronoi tessellation and the Delaunay triangulation, respectively.

3.4. Population Demographic History

Demographic history of *T. yarkandensis* in the Tarim river basin was reconstructed based on the Bayesian skyline plot using the skyline function in the “ape” package (Figure 5). The optimal estimated epsilon was 8.4882, and the estimate of effective population size through time showed that *T. yarkandensis* did not experience historical population bottlenecks, and that the historical population size of *T. yarkandensis* in the Tarim river basin experienced an expansion between 100 and 80 thousand years ago (with estimation of substitution rate = 0.0025). Then, the population size underwent a steady period. After that, the population size experienced a modest increase 40 thousand years ago.

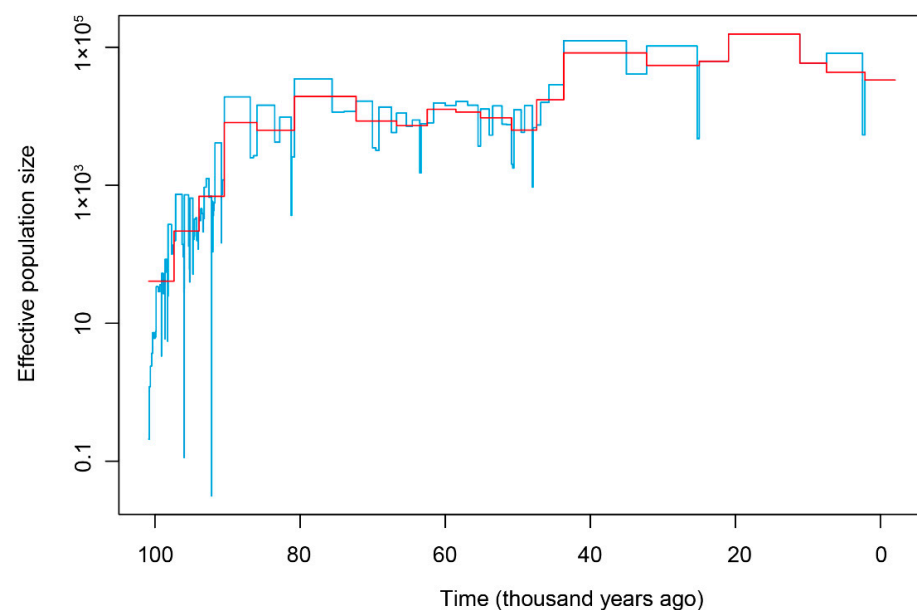


Figure 5. The Bayesian skyline plot of *T. yarkandensis*. The estimate of time unit was based on the substitution rate = 0.0025.

4. Discussion

The Tarim River basin is the main distribution area of *T. yarkandensis* natural populations. Comprehensive population genetic studies of the *T. yarkandensis* populations in the Tarim River basin contribute to revealing the genetic diversity of natural populations and evaluating the status of germplasm resource. Actually, most areas of the Tarim River basin belong to a depopulated zone with high altitude. Meanwhile, due to the river diversions and cutoffs of the Tarim River basin, the sample collection was quite difficult for biological researchers. In this study, we spent several months collecting the wild samples of *T. yarkandensis* for the population genetic study in 2019. Unfortunately, no *T. yarkandensis* individual was caught in some sampling locations; thus, several populations in the upper reaches of the Tarim River basin were absent in the present study. More populations from other locations (e.g., Bosten lake, Hotan River and Toxkan River) will be collected for systematic evaluation of genetic resource in our future studies.

Given the lack of whole-genome sequences of *T. yarkandensis*, we selected de novo mapping pipeline with Stacks for SNP identification. For genetic analyses, we used the populations parameter of “-write_single_snp” to trim the outlier SNPs, which could impact the accurateness of genetic structure and genetic differentiation inference. A total of 125,893 high-quality SNPs were detected and used for population genetic analyses in this study. To further eliminate the genetic bias, we filtered the obtained loci using the HWE test, and used the filtered loci to carry out the structure analyses and PCA. Overall, it could be concluded that the nine populations of *T. yarkandensis* exhibited high genetic diversity, and atypical genetic differentiation was observed among these populations. Particularly, the genetic variations of these populations mainly derived from the among-population variations. We found that the pairwise F_{ST} values among these populations were very high, indicating that the genetic exchange among these populations of *T. yarkandensis* was very limited compared with other small bottom-dwelling fish species, such as *M. anguillicaudatus* [14]. It could be possibly explained by the weak migration abilities in the wild and the fragments of habitats in the Tarim River basin.

According to the geographic locations, we assigned AKT, SWS, WC, JDT, WSX and TTM into south populations, which are distributed in the lower reaches of the Tarim River basin, and DWQ, TKX and KZE were assigned into north populations, which are located in the upper reaches of the Tarim River basin. Based on the results of population structure and the NJ tree, the nine populations were separated into two clusters, but the level of genetic differentiation between two clusters was moderate. Most individuals from south populations were clustered together, and the two clusters could not be completely separated based on the results of genetic admixture and phylogenetic analysis. Notably, most individuals from TTM showed a closer affinity to the north populations. TTM is located in Taitema Lake, which is the terminal lake of the main stream of the Tarim River and the beginning of the Qarqan River. Taitema Lake was dried up and had experienced progressive degradation without water flows over several decades until the ecological water conveyance project (EWCP) was initiated by transferring water from Bosten Lake to the terminal Taitema Lake [32]. We speculated that the implementation of EWCP also accelerated the gene flow from populations located in upper reaches of the Tarim River basin to the TTM population. Actually, aquaculture activities in the Tarim river basin are less developed due to the special environment (high salinity and sediment concentration) [33,34], and the natural populations of *T. yarkandensis* were actually less interfered by agricultural activities. In this context, although anthropogenic factors could affect the genetic structure and differentiation of fish populations [35,36], especially in some economic fish species widely cultured in aquaculture [14,37], we noted that the genetic exchange among most populations mediated by anthropogenic factors was limited in the Tarim River basin. Here, we suggest that the *T. yarkandensis* populations should be monitored in the long term to avoid significant reductions in abundance due to the genetic characterizations, including the high levels of genetic variation originating from among-population variation and the small effective population size. Most importantly,

these genetic characterizations together with the fragmented and isolated habits in the Tarim River basin increase the probability of outbreeding depression. The fine-scale genetic differences among populations should be regarded with caution in the conservation of *T. yarkandensis*. Multi-generational outbreeding depression can reduce fitness to a greater extent than reductions generated by inbreeding depression [38]. In this case, although high levels of genetic variation within populations were detected, we suggest joint evaluation of inbreeding and outbreeding risks in the wild in the future conservation and management of this species.

In the past decades, traditional techniques and strategies have been established to monitor biological diversity of fish populations in the wild. With the advance of genome-wide molecular markers, study of fish population genetics is increasingly important and popular in ecological, evolutionary and conservation fields. As mentioned above, *T. yarkandensis* is native to the Tarim River basin, and the wild capture of *T. yarkandensis* rarely occurred. However, biological invasion of other alien fish species in the Tarim River basin, such as *Paramisgurnus dabryanus*, *Perca fluviatilis*, and *Ophiocephalus argus*, severally impact the native fish biodiversity [39]. Actually, the rivers in the lower reaches of the Tarim River basin underwent cutoff in the past decades due to climate change and dam construction in the upper and middle reaches. From May 2000 to November 2001, local government initiated the implementation of EWCP by transferring water from Bosten Lake to the lower reaches of the Tarim River [40], which brings many positive direct effects, such as elevated groundwater level, vegetation coverage recovery and agricultural irrigation; however, this project still have some adverse effects, especially including alien aquatic species invasion. Those alien fish species can be well adapted to the habitats of the Tarim River basin and further result in the decline of native fish species by predation and competition. According to the results of a fisheries resource survey in 2019 (unpublished work), we found that the composition of fish communities in the lower reaches was numerically dominated by small fish species, such as *Pseudorasbora parva*, *Carassius auratus*, and *Hypomesus olidus*, which was different to the upper reaches. Therefore, scientific control of alien fish should be implemented to protect the native fish species in the Tarim River basin, especially in the lower reaches. Meanwhile, artificial breeding and releasing could be adopted as conservation and management approaches for *T. yarkandensis*. Importantly, with the releasing of artificially produced offspring into wild, the genetic consequences of these releases on natural populations should be evaluated to maintain genetic diversity and avoid degeneration of genetic resources in further studies. Overall, comprehensively investigating the genetic diversity and population structure contributed to understanding the genetic status of natural resources of *T. yarkandensis* in the Tarim River basin, and this study laid a foundation for the conservation and management of natural *T. yarkandensis* populations.

5. Conclusions

In this study, a total of 129,873 genome-wide SNPs of *T. yarkandensis* were obtained with the Stacks pipeline for population genetic analyses. High levels of genetic diversity were detected among nine populations in the Tarim River basin. The majority of genetic variations originated from among-population variations, and a high levels of genetic differentiation among these populations was observed. The nine populations could be separated into two clusters (i.e., south and north populations), and modest genetic differentiation between south and north populations was observed, while the individuals from several populations were not clustered together by geographical location. *T. yarkandensis* populations in the Tarim River basin had not experienced genetic bottlenecks, and the effective population size remained stable. This study provides useful genetic information for conservation and management of *T. yarkandensis* populations in the Tarim River basin.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/biology10080734/s1>, Table S1: The statistics of raw data generated by Illumina sequencing. Table S2: AMOVA result of nine *T. yarkandensis* populations.

Author Contributions: X.Z. and S.Y. conceived the study; X.Z., W.Z. extracted the DNA samples. Q.Z., J.S., D.L. and B.H. collected the samples. X.Z. and S.Y. analyzed the genetic data and wrote the manuscript; X.Z., R.T. and S.Y. revised the manuscript; and all authors approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially 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).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Scientific Ethic Committee of Huazhong Agricultural University (SYXK2020-0084; 2020-11-18).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this manuscript and its Supplementary Materials and are available on request from the corresponding authors.

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

References

- Chen, Y.; Xu, C.; Hao, X.; Li, W.; Chen, Y.; Zhu, C.; Ye, Z. Fifty-year climate change and its effect on annual runoff in the Tarim River Basin, China. *Quat. Int.* **2009**, *208*, 53–61.
- Chen, S.A.; Hou, J.; Yao, N.; Xie, C.; Li, D. Comparative transcriptome analysis of *Triplophysa yarkandensis* in response to salinity and alkalinity stress. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2019**, *33*, 100629. [[CrossRef](#)]
- Hou, F.; Zhang, X.; Zhang, X.; Yue, B.; Song, Z. High intra-population genetic variability and inter-population differentiation in a plateau specialized fish, *Triplophysa orientalis*. *Environ. Biol. Fishes* **2012**, *93*, 519–530. [[CrossRef](#)]
- Feng, C.; Zhou, W.; Tang, Y.; Gao, Y.; Chen, J.; Tong, C.; Liu, S.; Wanghe, K.; Zhao, K. Molecular systematics of the *Triplophysa robusta* (Cobitoidea) complex: Extensive gene flow in a depauperate lineage. *Mol. Phylogenet. Evol.* **2019**, *132*, 275–283. [[CrossRef](#)]
- Ren, Q.; Yang, J.-X.; Chen, X.-Y. Phylogeographical and Morphological Analyses of *Triplophysa stenura* (Cypriniformes: Nemacheilidae) from the Three Parallel Rivers Region, China. *Zool. Stud.* **2018**, *57*, e26.
- Nie, Z.; Hao, W.; Jie, W.; Xue, Z.; Ma, Z. Length-weight relationship and morphological studies in the Kashgarian loach *Triplophysa yarkandensis* (Day, 1877) from the Tarim River, Tarim River Basin, North-West China. *Indian J. Fish.* **2013**, *60*, 15–19.
- Yao, N.; Wang, S.; Chen, S.; Cheng, Y.; Zhang, Y.; Xie, C. Tolerance of *Triplophysa yarkandensis* Larvae to Salinity and Alkalinity. *Guizhou Agric. Sci.* **2016**, *44*, 117–120.
- Ning, X.; Zhang, Y.-Z.; Sui, Z.-H.; Quan, X.-Q.; Zhang, H.-G.; Liu, L.-X.; Han, Q.-D.; Liu, Y.-G. The complete mitochondrial DNA sequence of Kashgarian loach (*Triplophysa yarkandensis*) from Bosten Lake. *Mitochondrial DNA Part B* **2020**, *5*, 821–823. [[CrossRef](#)] [[PubMed](#)]
- Gong, X.; Xu, H.; Xiang, W.; Fan, Z.; Zheng, L.; Wang, C. Complete mitochondrial genome of Kashgarian loach, *Triplophysa yarkandensis* (Day, 1877) in the Tarim river. *Mitochondrial DNA Part A* **2016**, *27*, 3192–3193. [[CrossRef](#)]
- Sherman, K.D.; Paris, J.R.; King, R.A.; Moore, K.A.; Stevens, J.R. RAD-Seq Analysis and in situ Monitoring of Nassau Grouper Reveal Fine-Scale Population Structure and Origins of Aggregating Fish. *Front. Mar. Sci.* **2020**, *7*, 157. [[CrossRef](#)]
- Kang, J.; Ma, X.; He, S. Population genetics analysis of the Nuijiang catfish *Creteuchiloglanis macropterus* through a genome-wide single nucleotide polymorphisms resource generated by RAD-seq. *Sci. Rep.* **2017**, *7*, 2813. [[CrossRef](#)]
- Chen, K.-Y.; Euclide, P.T.; Ludsin, S.A.; Larson, W.A.; Sovic, M.G.; Gibbs, H.L.; Marschall, E.A. RAD-Seq Refines Previous Estimates of Genetic Structure in Lake Erie Walleye. *Trans. Am. Fish. Soc.* **2020**, *149*, 159–173. [[CrossRef](#)]
- Sunde, J.; Yildirim, Y.; Tibblin, P.; Forsman, A. Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (*Esox lucius*) and a Synthesis of Previous Studies. *Front. Genet.* **2020**, *11*, 218. [[CrossRef](#)]
- Yi, S.; Wang, W.; Zhou, X. Genomic evidence for the population genetic differentiation of *Misgurnus anguillicaudatus* in the Yangtze River basin of China. *Genomics* **2019**, *111*, 367–374. [[CrossRef](#)] [[PubMed](#)]
- Zhong, J.; Yi, S.; Ma, L.; Wang, W. Evolution and phylogeography analysis of diploid and polyploid *Misgurnus anguillicaudatus* populations across China. *Proc. R. Soc. B Biol. Sci.* **2019**, *286*, 20190076. [[CrossRef](#)] [[PubMed](#)]
- Yi, S.; Li, Y.; Shi, L.; Zhang, L.; Li, Q.; Chen, J. Characterization of Population Genetic Structure of red swamp crayfish, *Procambarus clarkii*, in China. *Sci. Rep.* **2018**, *8*, 5586. [[CrossRef](#)]
- Thongda, W.; Lewis, M.; Zhao, H.; Bowen, B.; Lutz-Carrillo, D.J.; Peoples, B.K.; Peatman, E. Species-diagnostic SNP markers for the black basses (*Micropterus* spp.): A new tool for black bass conservation and management. *Conserv. Genet. Resour.* **2020**, *12*, 319–328. [[CrossRef](#)]
- Takahashi, T.; Nagano, A.J.; Kawaguchi, L.; Onikura, N.; Nakajima, J.; Miyake, T.; Suzuki, N.; Kanoh, Y.; Tsuruta, T.; Tanimoto, T.; et al. A ddRAD-based population genetics and phylogenetics of an endangered freshwater fish from Japan. *Conserv. Genet.* **2020**, *21*, 641–652. [[CrossRef](#)]

19. Yang, X.; Liu, H.; Ma, Z.; Zou, Y.; Zou, M.; Mao, Y.; Li, X.; Wang, H.; Chen, T.; Wang, W.; et al. Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Mol. Ecol. Resour.* **2019**, *19*, 1027–1036. [[CrossRef](#)]
20. Sun, X.; Liu, D.; Zhang, X.; Li, W.; Liu, H.; Hong, W.; Jiang, C.; Ning, G.; Ma, C.; Zeng, H. SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PLoS ONE* **2013**, *8*, e58700. [[CrossRef](#)]
21. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
22. Rochette, N.C.; Rivera-Colon, A.G.; Catchen, J.M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **2019**, *28*, 4737–4754. [[CrossRef](#)]
23. Puritz, J.B.; Hollenbeck, C.M.; Gold, J.R. dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2014**, *2*, e431. [[CrossRef](#)]
24. Excoffier, L.; Lischer, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [[CrossRef](#)]
25. Fricot, E.; Francois, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **2015**, *6*, 925–929. [[CrossRef](#)]
26. Raj, A.; Stephens, M.; Pritchard, J.K. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **2014**, *197*, 573–589. [[CrossRef](#)] [[PubMed](#)]
27. Jombart, T.; Kamvar, Z.N.; Collins, C.; Lustrik, R.; Beugin, M.P.; Knaus, B.J.; Solymos, P.; Mikryukov, V.; Schliep, K.; Maié, T. Adegenet: Exploratory Analysis of Genetic and Genomic Data. *Doc. Ophthalmol.* **2018**, *107*, 13–36.
28. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.; Bender, D.; Maller, J.; Sklar, P.; Bakker, P.; Daly, M.J. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)] [[PubMed](#)]
29. Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc.* **2011**, *174*, 245–246. [[CrossRef](#)]
30. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)] [[PubMed](#)]
31. Manni, F.; Guérard, E.; Heyer, E. Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier’s algorithm. *Hum. Biol.* **2004**, *76*, 173–190. [[CrossRef](#)] [[PubMed](#)]
32. Hou, P.; Beeton, R.J.S.; Carter, R.W.; Dong, X.G.; Li, X. Response to environmental flows in the lower Tarim River, Xinjiang, China: Ground water. *J. Environ. Manag.* **2007**, *83*, 371–382. [[CrossRef](#)] [[PubMed](#)]
33. Wang, Y.; Zhou, X.; Bernard, E. Water environment carrying capacity in Bosten Lake basin. *J. Clean. Prod.* **2018**, *199*, 574–583. [[CrossRef](#)]
34. Mamat, Z.; Halik, U.; Aishan, T.; Aini, A.; Hepp, L.U. Ecological effect of the riparian ecosystem in the lower reaches of the Tarim River in northwest China. *PLoS ONE* **2019**, *14*, e0208462. [[CrossRef](#)] [[PubMed](#)]
35. Zhai, D.-D.; Li, W.-J.; Liu, H.-Z.; Cao, W.-X.; Gao, X. Genetic diversity and temporal changes of an endemic cyprinid fish species, *Ancherythroculter nigrocauda*, from the upper reaches of Yangtze River. *Zool. Res.* **2019**, *40*, 427–438. [[CrossRef](#)]
36. Valenzuela-Aguayo, F.; McCracken, G.R.; Manosalva, A.; Habit, E.; Ruzzante, D.E. Human-induced habitat fragmentation effects on connectivity, diversity, and population persistence of an endemic fish, *Percilia irwini*, in the Biobío River basin (Chile). *Evol. Appl.* **2020**, *13*, 794–807. [[CrossRef](#)]
37. Soliman, T.; Aly, W.; Fahim, R.M.; Berumen, M.L.; Jenke-Kodama, H.; Bernardi, G. Comparative population genetic structure of redbelly tilapia (*Coptodon zillii* (Gervais, 1848)) from three different aquatic habitats in Egypt. *Ecol. Evol.* **2017**, *7*, 11092–11099. [[CrossRef](#)]
38. Houde, A.L.; Fraser, D.J.; O’Reilly, P.; Hutchings, J.A. Relative risks of inbreeding and outbreeding depression in the wild in endangered salmon. *Evol. Appl.* **2011**, *4*, 634–647. [[CrossRef](#)]
39. Chen, G.; Qiu, Y.; Li, L. Fish invasions and changes in the fish fauna of the Tarim Basin. *Acta Ecol. Sin.* **2017**, *37*, 700–714.
40. Ye, Z.; Chen, Y.; Li, W.; Yan, Y. Effect of the ecological water conveyance project on environment in the Lower Tarim River, Xinjiang, China. *Environ. Monit. Assess.* **2009**, *149*, 9–17. [[CrossRef](#)]