



# Article Genetic Diversity and Signatures of Selection in the Roughskin Sculpin (*Trachidermus fasciatus*) Revealed by Whole Genome Sequencing

Lize San <sup>1,2</sup>, Zhongwei He <sup>1,2</sup>, Yufeng Liu <sup>1,2</sup>, Yitong Zhang <sup>1,2</sup>, Wei Cao <sup>1,2</sup>, Jiangong Ren <sup>1,2</sup>, Tian Han <sup>1,2,3</sup>, Bingbu Li <sup>1,2</sup>, Guixing Wang <sup>1,2</sup>, Yufen Wang <sup>1,2</sup> and Jilun Hou <sup>1,2,\*</sup>

- $^1$   $\;$  Hebei Key Laboratory of the Bohai Sea Fish Germplasm Resources Conservation and Utilization,
- Beidaihe Central Experiment Station, Chinese Academy of Fishery Sciences, Qinhuangdao 066100, China
- <sup>2</sup> Bohai Sea Fishery Research Center, Chinese Academy of Fishery Science, Qinhuangdao 066100, China
- <sup>3</sup> Ocean College, Hebei Agricultural University, Qinhuangdao 066009, China
- \* Correspondence: houjl@bces.ac.cn

**Simple Summary:** The roughskin sculpin (*Trachidermus fasciatus*) was previously a locally abundant species in China, but the habitat required for the survival and breeding of this species has been seriously damaged. Its resources have been almost depleted, and it has been listed as a national second-class protected animal. In recent years, the population of roughskin sculpin has recovered in the natural environment through enhancement programs and the release of juveniles. However, the effects of released roughskin sculpin on wild populations' genetic structure and diversity remain unclear. In this study, the whole-genome resequencing of two cultured populations and one wild population showed that the two cultured populations have similar genetic diversity compared with the wild population. In addition, artificial feeding affected the formation of traits of roughskin sculpin. The finding provides data support for the resource recovery of roughskin sculpin.

**Abstract:** The roughskin sculpin (*Trachidermus fasciatus*) is an endangered fish species in China. In recent years, artificial breeding technology has made significant progress, and the population of roughskin sculpin has recovered in the natural environment through enhancement programs and the release of juveniles. However, the effects of released roughskin sculpin on the genetic structure and diversity of wild populations remain unclear. Studies on genetic diversity analysis based on different types and numbers of molecular markers have yielded inconsistent results. In this study, we obtained 2,610,157 high-quality SNPs and 494,698 InDels through whole-genome resequencing of two farmed populations and one wild population. Both farmed populations showed consistent levels of genomic polymorphism and a slight increase in linkage compared with wild population, but the degree of genetic differentiation was low (overall average Fst = 0.015). Selective sweep analysis showed that 523,529 genes were selected in the two farmed populations, and KEGG enrichment analysis showed that the selected genes were related to amino acid metabolism, which might be caused by artificial feeding. The findings of this study provide valuable additions to the existing genomic resources to help conserve roughskin sculpin populations.

Keywords: Trachidermus fasciatus; genetic diversity; population structure; selective sweeps

## 1. Introduction

The roughskin sculpin (*Trachidermus fasciatus*), belonging to the Cottidae family, is a kind of small, bottom-dwelling catadromous migrating fish mainly distributed in the Gulf of Southern Japan, the west coast of the Korean Peninsula, the northeast coast of China, and the downstream of their corresponding rivers [1]. China's roughskin sculpin was previously a locally abundant species, but due to the aggravation of environmental pollution in the



Citation: San, L.; He, Z.; Liu, Y.; Zhang, Y.; Cao, W.; Ren, J.; Han, T.; Li, B.; Wang, G.; Wang, Y.; et al. Genetic Diversity and Signatures of Selection in the Roughskin Sculpin (*Trachidermus fasciatus*) Revealed by Whole Genome Sequencing. *Biology* 2023, 12, 1427. https://doi.org/ 10.3390/biology12111427

Academic Editor: Jose Pedro Andrade

Received: 24 September 2023 Revised: 27 October 2023 Accepted: 10 November 2023 Published: 13 November 2023



**Copyright:** © 2023 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/). surrounding seas, coupled with overfishing and the construction of dams, the habitat required for the survival and breeding of this species has been seriously damaged. Its resources have been almost depleted, and it has been listed as a national second-class protected animal [2,3]. Currently, destruction of the habitat of the roughskin sculpin in China accounts for the significant difficulties faced with respect to its research and protection. It is well established that when the population is small, species are easily affected by genetic drift, inbreeding depression, and loss of adaptive genetic variation, resulting in decreased population genetic diversity [4]. The decrease in the population's genetic diversity affects its stability and fitness and ultimately increases the risk of extinction. Significant inroads have been achieved over the past few years in artificial breeding and mass production of seeds of roughskin sculpin, which is of great significance to protecting the species' resources and maintaining its population size [5,6].

Elucidating the genetic diversity and genetic structure of the roughskin sculpin population is crucial for formulating conservation measures. However, there are inconsistencies in the current research on the genetic differentiation of the roughskin sculpin population. Indeed, it remains unclear whether there is a common genetic structure among Chinese roughskin sculpin populations [7]. This controversy may exist because a higher resolution than that provided by currently used molecular markers is needed, and the coverage and detection level of different molecular markers affect the reliability of the analysis of the genetic structure of the population. Over the years, random amplification of polymorphic DNA (RAPD), inter simple sequence repeat (ISSR), simple sequence repeat (SSR), mitochondrial DNA, and other molecular marker techniques have been used to study the genetic diversity and genetic differentiation levels of different populations of roughskin sculpin but yielded heterogeneous results [8–10]. Microsatellite markers have been used to analyze the genetic diversity of roughskin sculpin from the two geographical groups of the Yalu River and the Yellow River. It was found that the roughskin sculpin of both geographical groups had high levels genetic diversity, and the genetic diversity of the Yalu River population was slightly higher than that of the Yellow River population [11]. The microsatellite markers were used to analyze the genetic diversity of roughskin sculpin from eight geographical groups in Japan and China. The results showed that the wild roughskin sculpin had high levels of genetic diversity (Ho: 0.7~0.9) and genetic structure [12]. However, the number of microsatellite markers for a single test is limited, and microsatellites can only detect simple variations with low sensitivity, which is not conducive to other population genetic analyses, such as the study of interpopulation selection signals.

Population genomics have developed rapidly with the availability of high-throughput sequencing and the reduced costs of related technologies. Population genomics can screen thousands of genetic markers simultaneously at the genome scale, resulting in tremendous innovations in population genetics, conservation genetics, ecology, and evolutionary research [13–15]. For species whose genomes have been assembled, resequencing individuals of different populations based on high-throughput sequencing technology can obtain genomic information of different populations quickly and efficiently. In fish, genome-wide resequencing analyzes fish population structure and the history of population evolution by screening candidate genes associated with genomic regions and elucidating the molecular mechanisms underlying various genetic traits. In this respect, Xu et al., resequenced 33 representative carp individuals worldwide using genome-wide data, and the results showed that carp originated from two subspecies. A total of 326 candidate genes were obtained by selecting and eliminating the genes associated with carp scale deletion and red body color, which laid a foundation for revealing the molecular mechanism of trait differences in different carp (Cyprinus carpio) populations [16]. In a similar study of Atlantic salmon (Salmo salar L.), the Atlantic Ocean's eight populations in northern and southern Norway were analyzed using genome resequencing, revealing that natural selection acting on immune-related genes led to genetic differences between populations in the Norwegian Atlantic and that the plasticity of genomes may promote differences between populations [17]. Population genetic studies conducted using molecular markers represent

a necessary technical means to understand population adaptive evolution. In this respect, enrichment analysis of selected regions revealed that the expanded gene family of silver carp (*Hypophthalmichthys molitrix*) was associated with disease, immune system, and environmental adaptation [18]. Whole-genome resequencing technology can also be used to obtain the genomic information of each subpopulation of a species' natural population, and a large amount of various information can be obtained. Through the analysis of genetic variations, various biological issues, such as population genetic structure, gene exchange, speciation mechanism, and population evolution dynamics, can be explored. This can be illustrated briefly by the evolutionary origin and domestication history of studying goldfish (*C. auratus*). With genome-wide variation information, genes associated with morphogenesis, pigmentation, behavior, immune response, infectious disease, energy metabolism, and hormone response have been successfully mapped [19].

The reference genome of roughskin sculpin has been precisely assembled, laying the groundwork for subsequent population genetic analysis and resource conservation [20]. In this study, we obtained the genome-wide SNP of roughskin sculpin through resequencing. The genetic diversity of roughskin sculpin in farmed and wild populations was analyzed based on genomic SNP, and the influence of selection on the genome of roughskin sculpin under captivity conditions was explored, guiding species conservation through artificial release.

## 2. Materials and Methods

## 2.1. Ethics

This study was performed in accordance with the Guidelines for Care and Use of Laboratory Animals provided by the Chinese Association for Laboratory Animal Sciences (No. 2011–2). The study protocols were approved by the Animal Care and Use Committee of Beidaihe Central Experiment Station, Chinese Academy of Fishery Sciences (CAFS).

### 2.2. Sample Collection and Measurement of Growth Traits

A total of 192 T. fasciatus were collected from three different geographic populations (Qinhuangdao City of Hebei province (n = 64, weight:  $30.66 \pm 10.4$  g), Tianjin Municipality (n = 64, weight: 40.46  $\pm$  9.84 g), and Dongying City of Shandong province (n = 64, weight:  $30.47 \pm 7.21$  g), China). The detailed geographical locations of the sampling are shown in Figure 1. According to the origin of the population, they were named QHD, TJ, and SD. The SD (temperature:  $12 \pm 1$  °C, DO:  $6.5 \pm 0.5$  mg/L, pH:  $7.5 \pm 0.5$ , salinity:  $24 \pm 1$ %) and TJ groups (temperature:  $11 \pm 1$  °C, DO:  $6.5 \pm 0.5$  mg/L, pH:  $7.5 \pm 0.5$ , salinity:  $24 \pm 1$ %) were cultured, and QHD (temperature:  $10 \pm 1$  °C, DO:  $8.5 \pm 0.5$  mg/L, pH:  $8.0 \pm 0.5$ , salinity:  $31 \pm 1\%$ ) was a wild population. The breeding group was taken from the farm, and the sampling time was November 2022. Wild populations were caught by bottom trawling. The fishing site was in the offshore area of the Bohai Sea, the fishing depth range was 4~6 m, the fishing route was 30 km, and the fishing time was November 2022. Their morphometric parameters were measured, including body weight, height, and length, as well as total length. Four morphological traits of (body weight (BW), body length (BL), body height (BH), and total length (TL)) were measured after anesthesia with 30 g/m<sup>3</sup> MS-222. The weight was measured using and electronic balance with an accuracy of 0.1 g. The body length and height were measured with a vernier caliper, which is accurate to 0.1 cm. The measurement methods of the total length, body length, and body height of the sculpin are described as follows.



Figure 1. Map of sampling sites for roughskin sculpin.

- 1. Total length: the vertical distance from the front end of the maxillary kiss to the end of the caudal fin;
- 2. body length: the vertical distance from the front end of the maxillary kiss to the base of the caudal fin;
- 3. body height: the vertical distance at the highest point of the fish body.

SPSS 17.0 was used to test the normality of all measured body weight, total length, body length and body height data, and variance analysis and multiple comparisons were performed.

Thirteen fish were randomly selected in equal amounts from SD and TJ, and 14 were selected from QHD. The fin tissue of these 40 fish was collected, and DNA was extracted for sequencing.

## 2.3. DNA Isolation and Sequencing

Genomic DNA was extracted using a TIANamp marine animal DNA Extraction kit (TIANGEN, Beijing, China) according to the kit protocol. DNA quality was checked by 1% agarose gel electrophoresis and a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) spectrophotometer. Qualified DNA samples were randomly interrupted by a Covaris crusher (Covaris, Woburn, MA, USA) with a length of 350 bp. A TruSeq Library Construction Kit (Illumina, San Diego, CA, USA) was used for library construction according to the manufacturer's instructions. End repair, polyadenylation, the addition of a sequencing adapter, purification, and PCR amplification were conducted for the library preparation of DNA fragments. Illumina NGS technology was used to sequence the constructed libraries. After the library was constructed, Qubit3.0 was used for preliminary quantification, diluted to  $1 \text{ ng}/\mu\text{L}$ ; then, an Agilent 2100 was used to detect the insert size of the library. After the insert size met the requirements, the effective concentration of the library was accurately quantified by qPCR (effective concentration of the library > 2 nM) to ensure the quality of the library. After the library inspection was qualified, PE150 paired-end sequencing was performed on the Illumina platform according to the effective concentration of the library and data output requirements.

## 2.4. Variant Discovery and Quality Control

The raw data obtained through sequencing were filtered to remove low-quality sequencing data using fastp v0.20.0 software [21]. The filtering standards were as follows: (1) Reads containing Adapter sequences were filtered. (2) Reads with more than 10% of the total number of bases in a single end read that could not be determined were filtered. (3) Reads with low-quality bases exceeding 50% of the read length were filtered. Effective, high-quality sequencing data were aligned to the reference genome by BWA (--mem) [22] software, and the alignment results were used to remove duplicates (rmdup parameter) with SAMTOOLS [23]. SAMTOOLS, BCFTOOLS (filter -e 'QUAL < 10 || DP < 2') [24] software was used for population SNP detection. The polymorphism sites in the population were detected using the Eyes model, and high-quality SNPs were obtained by filtering and screening as follows: (1) SNPs with mass values below Q20 were filtered out. (2) SNPs < 5 bp apart were filtered out. (3) SNPS with a coverage depth below 2 times the average were filtered out. Finally, SnpEff [25] software was used to apply gene-based annotation, region-based annotation, filter-based annotation, and other functionalities to SNP detection results.

## 2.5. Population Genetic Diversity Analysis

Based on the filtered SNP variation file, PLINK (--het; --hardy) software [26] calculated the observed heterozygosity (Ho) and expected heterozygosity (He) of different population variation sites. The population fixation index (Fst) and nucleotide diversity ( $\theta\pi$ ) were calculated using VCFtools (--window-pi 50,000 --window-pi-step 10,000; --fst-window-size 50,000 --fst-window-step 10,000) [27], with the sliding window size set to 50 Kb and the window step set to 10 Kb. The strong selection signal within the selected region between the groups was obtained and visualized by drawing scatter plots in R. The genome-wide heterozygous rate (number of heterozygous SNPs/genome size) was calculated.

## 2.6. Linkage Disequilibrium (LD) and Population Structure Analysis

The correlation coefficient ( $r^2$ ) measures the degree of LD between different SNPs. The observed pair-wise LD ( $r^2$ ) was calculated using PopLDdecay [28] software (-MaxDist 500). The population structure was studied using principal component analysis (PCA) and population structure analysis of three populations. Using PLINK (v1.90) (--pca 20) software, individuals were clustered into subpopulations according to principal components based on the degree of difference in SNPs, and redundant SNPs with strong linkage were removed. The PLINK -indep-pairwise parameter was set to 100 20 0.5. Finally, individual ancestries were estimated using the model-based maximum likelihood method implemented in ADMIXTURE [29], which utilizes a fast numerical optimization algorithm to calculate estimates. Assuming that the population is composed of several subgroups (k = x), the simulation algorithm was used to find the most reasonable sample classification method in the case of k = x. Then, according to the maximum likelihood value of each simulation, we identified the k value that is most suitable for this population.

### 2.7. Screening for Selective Sweeps

The Fst and  $\pi$  ratio method was used to analyze the selection signals of different populations in pairs, and the intersection extracted from the value range of the top 5% was a significant selected region. The genes contained in these regions were the selected genes. The selected region was extracted using BEDtools software and genetically annotated based on the annotation GFF file.

## 2.8. KEGG Enrichment Analysis of Candidate Genes

KofamKOALA [30] software was used to perform KEGG enrichment analysis of *T. fasciatus* genes based on the HMM model, and a database for KEGG enrichment analysis was constructed. The clusterprofiler [31] R package was used for KEGG analysis of the selected genes. The resulting p value was corrected using the Benjamini–Hochberg (BH)

method to obtain the adjusted *p* value ( $P_{adj}$ ), and a  $P_{adj}$  value < 0.05 was considered statistically significant.

#### 3. Results

## 3.1. Comparison of Growth Traits in Different Geographic Groups

The differences in weight, body length, body height, and total length indicators between different populations are shown in Figure 2. The weight, body length, body height, and total length of individuals in the TJ population were significantly higher than in the QHD and SD populations. The body length, height, and total length of individuals in the QHD population were significantly higher than in the SD population, but there was no significant difference in weight between the QHD and SD populations. The biometric data (minimum, maximum, average and standard error) of different populations of roughskin sculpin are shown in the Table S1. Interestingly, we found that compared with the wild population in QHD, the SD population, as an artificial breeding population, experienced no increase in morphometric parameters. Overall, the morphometric parameters of the TJ population, which was also subject to artificial breeding, was significantly superior to those of the QHD population.



**Figure 2.** Box plot of morphometric statistics for growth indices of different populations of roughskin sculpin. The number of samples measured in each population was 64. (**A**) Weight; (**B**) body height; (**C**) body length; (**D**) total length.

## 3.2. SNP Characteristics

All 40 fishes were genotyped for 2,610,157 high-quality SNPs and 494,698 InDels in this study. The variants were filtered with a call rate > 90% and MAF > 0.01 in each population. SnpEff annotation revealed that 52.8% of variants were located in intergenic regions, and 47.2% were located in gene regions, including exons, introns, and untranslated regions. The variants on introns accounted predominantly for variants located in the gene region (41.4%).

Only 6.0% of variants were located in the exon region, containing 152,974 synonymous and 276,614 nonsynonymous substitutions (Table 1).

Category	Туре	Count	Percent
Region	Intergenic	3,331,449	47.0%
	Intron	2,985,075	41.4%
	Exon	429,588	6.0%
	5′ UTR	125,011	3.9%
	3' UTR	280,506	1.7%
Mutation	Synonymous	152,974	2.1%
	Non-synonymous	276,614	3.9%

Table 1. The number of variant effects by region and type.

#### 3.3. Genetic Diversity

Genetic diversity among populations was assessed by estimating observed and expected heterozygosity, genetic distance, and nucleotide diversity, as shown in Table 2. The observed heterozygosity (0.3524~0.3549) and expected heterozygosity (0.3343~0.3346) of the three groups did not differ significantly. The observed heterozygosity was not significantly different from the expected heterozygosity. The proportions of the polymorphic SNPs were 98.39%, 98.89%, and 98.89% in the wild and two farmed populations, respectively. The proportions of the polymorphic SNPs in the two farmed populations were similar but higher than in the wild population. Genetic distance (D) was used to measure the degree of genetic difference among samples within a population. The approximate genetic distance for the QHD, SD, and TJ was 0.274212, 0.278117, and 0.277829, respectively.

Table 2. Summary of genetic diversity measurement for three populations.

Statistic	QHD	SD	TJ
Но	0.3524	0.3539	0.3549
He	0.3344	0.3343	0.3346
MAF	0.2437	0.2432	0.2434
P <sub>N</sub>	98.39%	98.89%	98.89%
D	0.274212	0.278117	0.277829

### 3.4. Population Differentiation and Linkage Disequilibrium

By analyzing the degree of genetic differentiation among the three populations, it was found that the QHD population exhibited the highest genetic differentiation from the SD and TJ populations, with a mean Fst of 0.02. The SD population had the lowest genetic differentiation compared with the TJ population, with a mean Fst of 0.005 (Figure 3). Because the Fst values of the three populations were all less than 0.05, we considered that the genetic composition of the three populations was similar and that the degree of genetic differentiation was low. Genome-wide linkage imbalance (LD) analysis showed that the linkage degree of SPNs in the three groups was low (0~0.15) (Figure 4). The decay of the LD coefficient ( $r^2$ ) was slower in the QHD population than in the SD and TJ populations, which were cultured and subjected to a certain degree of selection. In addition, we found that the  $r^2$  attenuation decay curves of the SD and TJ populations overlapped, indicating that the two populations exhibited a low degree of differentiation, consistent with the Fst results.



**Figure 3.** Manhattan plots of Fst values. The X axis represents different chromosome names, and the Y axis represents the Fst values within the corresponding chromosome window. The different colors of point represent different chromosomes.



**Figure 4.** Genome-wide linkage disequilibrium (LD) decay plots. The x axis represents the distance of LD, and the y axis represents the linkage disequilibrium correlation coefficient.

Principal component analyses were performed on the QHD, SD, and TJ populations to explore their population structures. The first three were selected as clustering factors according to the variance-explained rate of principal component factors. The QHD population could be clearly distinguished from the SD and TJ populations by three principal components (Figure 5). The SD population overlapped with the TJ population. During the ADMIXTURE analysis, none of the k values' cross-validations (CVs) converged. In this study, we also determined the optimal number of subgroups (k = 3) by taking the actual number of geographical populations and dividing those three populations into two (Figure 6). The SD and TJ populations were clustered into one population, consistent with the Fst and principal component analysis results.



**Figure 5.** Sample structure indicated by principal component analysis (PCA) with the first three principal components.



**Figure 6.** Admixture structure plot. Each column represents an individual, where the length of the different colored segments indicates the proportion of the individual that belongs to different groups.

## 3.5. Selective Sweep and Selected Gene Annotation

Combined with Fst and the ratio of PI between different populations, strong selection signals were detected, and target genes were screened. During the comparisons of QHD and SD, QHD and TJ, and SD and TJ populations, we took the highest 5% of Fst and the window with the highest 5% (high top5%) and the lowest 5% (low top5%) Pi ratios as potential selection regions and merged adjacent and overlapping windows (Figures 7A, 8A and 9A). We used QHD as the reference population, since it is a wild population. At a significance level of 5%, 709 selective sweep regions were found in the TJ population compared to the QHD population. Approximately 35.45 MB and 523 genes were detected in the selective region. Furthermore, 629 selective sweep regions were found in the SD population compared to the QHD population, and approximately 31.45 MB and 529 genes were detected in the selective region. A previous population structure analysis showed that the QHD and TJ populations were closely related. Accordingly, the selected genes of the two groups were compared. It was found that 44.93% of genes between the two populations were homologous. Compared to SD, the TJ population had 457 selected regions containing 413 genes.



**Figure 7.** Identification of the domestication-selective sweeps and the KEGG enrichment analysis. (**A**) The  $\theta\pi$  ratio and Fst values of QHD and SD (green points indicate selection; black points represent not selection.). Line represents the selection threshold of 0.05. (**B**) The KEGG enrichment analysis of selected genes of SD.

The annotated genes underwent KEGG enrichment analysis to narrow the selected gene range. A total of 300 of the 500 genes in the TJ population were annotated to the corresponding KOs, and enrichment analysis showed that these genes were enriched in apoptosis, phenylalanine metabolism, and other pathways, but no significant functional clusters were identified during KEGG pathway enrichment analysis (Figure 7B). The enrichment analysis revealed several genes linked to the autophagy-animal, toll-like receptor signaling pathway, phenylalanine metabolism, ECM–receptor interaction, the Wnt signaling pathway, and other pathways. Enrichment analysis of selected genes in the SD population showed similar enrichment results to the TJ population, regardless of the significance level or the enrichment pathway. Besides pathways enriched in the TJ population, other pathways enriched in the SD population included cell adhesion molecules (Figure 8B). KEGG enrichment analysis of selected genes in the SD

population showed that the results differed from those with the QHD population as the reference population. The selected genes were associated with other pathways, such as cytokine–cytokine receptor interaction, cytokine receptors, and PPAR signaling pathways, although no statistical significance was found (Figure 9B).



**Figure 8.** Identification of the domestication-selective sweeps and the KEGG enrichment analysis. (**A**) The  $\theta\pi$  ratio and Fst values of QHD and TJ (green points indicate selection; black points represent not selection.). Line represents the selection threshold of 0.05. (**B**) The KEGG enrichment analysis of selected genes of TJ.



**Figure 9.** Identification of the domestication-selective sweeps and the KEGG enrichment analysis. (**A**) The  $\theta\pi$  ratio and Fst values of SD and TJ (green points indicate selection; black points represent not selection.). Line represents the selection threshold of 0.05. (**B**) The KEGG enrichment analysis of selected genes of TJ.

## 4. Discussion

The roughskin sculpin (*T. fasciatus*) was once widely distributed in the Northwest Pacific, but its habitat has been fragmented due to human activities such as dam construction and environmental pollution, resulting in a significant decrease in its population. Nowadays, it is only found in certain river basins of the Yellow and Bohai Seas and the East China Sea in China. The endangered status of the roughskin sculpin has prompted the implementation of conservation measures, such as stocking and releasing. Research on artificial breeding technology of the roughskin sculpin, including the production and aquaculture of the species, has made breakthroughs that not only protect the population of the roughskin sculpin but also meet market demand. Due to differences in breeding methods, growth patterns, feed sources, and feeding frequencies, there are differences in growth characteristics between wild and cultured populations [32,33]. In the natural sea area, roughskin sculpin mainly eat small shrimps and small fish, but most of Songjiang bass eat small shrimps. Palaemonetes sinensis sollaud and Caridina nilotica gracilipes de Man are the main shrimps ingested by roughskin sculpin. The breeding population of roughskin sculpin is fed commercial compound feed. The main source of protein in commercial feed is fish meal. Given that the cultured populations are fed on a fixed basis and have access to sufficient food during the breeding process, the morphological characteristics of the cultured population are significantly better than those of the wild population. Therefore, morphological characteristics represent important and stable indicators for distinguishing sea bream and sea bass from other marine fish [34–36]. In this study, the two cultured populations showed different morphological characteristics than wild populations. The three populations were simultaneously assessed to avoid the effect of confounders such as growth time. The weighted index of the SD population was similar to that of the QHD wild population, but the length, height, and other indices were significantly lower than those of the QHD population, which may be attributed to the fact that the artificial synthetic feed provided to the SD population could not meet the nutritional needs of the roughskin sculpin. Accordingly, the morphological characteristics of the TJ population, which is also a cultured population, were significantly better than those of the wild population. Moreover, it should be borne in mind that the survival environment leads to differences in body shape. Wild roughskin sculpin may require a longer body shape to adapt to the natural survival environment.

Population genetic diversity refers to species' genetic diversity within an area. Higher genetic diversity means that a species is more resilient to its environment. Current evidence suggests that changes in population size, structure, genetic mutations, and other factors can affect the population's genetic diversity. A study by Dewoody and Avise showed that the average heterozygosity of marine fish microsatellites was 0.79 and that the average heterozygosity of anadromous fish was 0.68 [37]. Li et al. revealed that the average observed heterozygosity of microsatellites in the roughskin sculpin population (0.83) was higher than the average of marine fish and higher than other Catadromous migration fish, such as the Japanese eel and the European eel [12,38,39]. Wild marine fish species Pangasianodon Hypophthalmus and Oreochromis niloticus have previously shown genomewide SNP heterozygosity of 0.04–0.60. In this study, the SNP heterozygosity of the wild population was found to be 0.335, which is comparable to that of the wild O. niloticus population [40,41]. However, Li et al. showed that the average heterozygosity of SNP in wild roughskin sculpin was 0.121, unlike the results of this study [7]. We speculate that this is due to the number of SNPs used for analysis—2,610,157 SNPs were used in this study to calculate heterozygosity, which is higher than that reported in Li et al.'s study (n = 10,153). In this respect, the population heterozygosity calculated by microsatellites was higher than that calculated by SNPs, suggesting that the number of markers affects heterozygosity. Theoretically, when the species range is limited and its effective population is decreased, its genetic diversity generally decreases due to genetic drift and inbreeding decline [42]. However, in this study, roughskin sculpin maintained a high level of genetic diversity, possibly due to its population size, which mitigated the effect of genetic drift. Higher levels

of genetic diversity have also been found in populations of some endangered species [43–45]. Affected by genetic drift and inbreeding, the genetic diversity of farmed populations was lower than that of wild populations. In the present study, we found that the heterozygosity of farmed and wild roughskin sculpin was similar. The roughskin sculpin is a fish with a life cycle of only one year, and both male and female fish die after spawning [46]. Therefore, the age structure of the roughskin sculpin population in the natural environment is relatively simple, avoiding an excessively close relationship between the collected samples caused by differences in generations. For the farmed population, the eggs of roughskin sculpin were bred by catching wild roughskin sculpin as parents; each broodstock can only be used for one year, which indirectly slows down the phenomenon of source degradation of the breeding population [47]. Overall, the above reasons account for the absence of significant differences among genetic diversity indicators observed in this study, such as heterozygosity, the proportions of the polymorphic SNPs, and genetic distance between wild and farmed roughskin sculpin populations. The main goals of conservation genetics is to protect the genetic diversity of species, improve the environmental adaptability of species, and reduce the risk of extinction. According to the results of this study, the genetic diversity of the wild population and that of the cultured population of roughskin sculpin are similar. In other words, it is feasible to restore the population size of roughskin sculpin in the natural sea area by supplementing with artificially bred roughskin sculpin. Moreover, the detection of genetic diversity of cultured populations before release can be used as a standard for stocking. The genetic diversity of roughskin sculpin in the natural sea area was detected regularly, and the effect of stocking was evaluated to a certain extent.

Population structure analysis results based on the genome-wide SNPs showed the genetic structure between the populations of roughskin sculpin. Principal component analysis showed that the three geographical groups of roughskin sculpin could be divided into two genetic groups: the SD and TJ population vs. the QHD population. Population structure analysis further verified the results of the principal component analysis. Although the cross-validation values of the predicted optimal population number did not converge, at K = 3, most individuals in the SD population and the TJ population were divided into one group, and the QHD population was independent of the other group. The SD population sampling point was far away from the TJ population. Given that there is theoretically no large-scale gene exchange, we speculate that the two groups come from the same spawning ground. Many molecular markers can be obtained by genome-wide testing, and due to their wide distribution in the genome, the fine genetic structure differences between populations can be detected. Compared with the QHD population, the differentiation coefficients of the SD and TJ populations were 0.0156 and 0.0158, respectively, with weak genetic differentiation. However, the SD and TJ populations did not show apparent genetic differentiation (Fst = 0.0039), consistent with the population structure analysis and principal component analysis results. Consistently, in Li's study, the genetic differentiation of wild roughskin sculpin from different geographical groups was not apparent [7].

Fst and  $\Pi$  are effective methods for detecting selective elimination regions, especially when mining functional areas closely related to the living environment, which can yield strong selection signals [48]. Indeed, the living environment, food source, composition, and other aspects of the breeding population differ completely from those of the wild population. Under the influence of human beings, the breeding population's relevant traits can be selected to better adapt to the breeding environment. Selective sweep analysis showed that captivity caused significant changes in the genome of roughskin sculpin. The upstream of the gene encoding L-amino acid oxidase (LAAO) was selected by comparing two cultured populations and one wild population. LAAO is an important oxidoreductase involved in the oxidative metabolism of amino acids in organisms. It can catalyze the oxidative deamination of L-amino acids with oxygen molecules as electron acceptors to produce corresponding keto acids, ammonia (NH<sub>3</sub>), and hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), which involve the metabolic pathways of tryptophan, phenylalanine, and tyrosine. Because the protein source of artificial compound feed is different from that of natural feed, under longterm regular feeding and management, the selected genes related to amino acid metabolism may play an important role in the process of metabolizing artificial compound feed. It has been established that a higher standard metabolic rate can lead to increased efficiency in food intake needed for growth. In this respect, a previous study comparing transcriptomic profiles of farmed and wild Atlantic salmon showed that differentially expressed genes were highly correlated with energy metabolism [49].

## 5. Conclusions

Overall, 2,610,157 high-quality SNPs and 494,698 InDels were identified using wholegenome sequencing, and the genetic diversity and population structure of two farmed and one wild population of roughskin sculpin were determined. The levels of genetic diversity were similar in all populations. Analysis of Fst values indicated low levels of genetic differentiation between the two farmed populations and wild populations and the lowest levels of genetic differentiation between the two farmed populations. According to population structure and principal component analysis, two clusters were detected in all the populations, and the two farmed populations could be clustered into one population, suggesting seedling exchange between the Shandong and Tianjin populations, which might affect the adaptability of roughskin sculpin. Selective sweep analysis showed that the selected genes in the two populations were related to amino acid metabolism, which might be caused by artificial feeding. These results provide essential information for the conservation of genetic diversity in roughskin sculping. Accordingly, corresponding genetic diversity conservation measures can be formulated based on to the genotyping results. Finally, this study demonstrates that whole-genome resequencing is an efficient genomewide SNP discovery tool that can accurately evaluate population genetics in roughskin sculpin.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biology12111427/s1, Table S1: The morphometric parameters data of different populations.

**Author Contributions:** Data curation, W.C. and B.L.; Formal analysis, L.S.; Funding acquisition, J.H.; Investigation, Z.H., Y.L. and J.R.; Methodology, Y.Z. and T.H.; Project administration, J.H.; Resources, Z.H., Y.L. and J.R.; Software, G.W.; Supervision, G.W. and Y.W.; Writing—original draft, L.S.; Writing—review and editing, J.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Central Public-Interest Scientific Institution Basal Research Fund, CAFS (2023TD41); the Key R&D Program of Hebei Province, China (21326307D); and the National Marine Genetic Resource Center.

**Institutional Review Board Statement:** This study was performed in accordance with the Guidelines for Care and Use of Laboratory Animals provided by the Chinese Association for Laboratory Animal Sciences (No. 2011–2). The study protocols were approved by the Animal Care and Use Committee of Beidaihe Central Experiment Station, Chinese Academy of Fishery Sciences (CAFS).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Raw DNA sequencing reads were deposited in the NCBI under project accession No. PRJNA945830.

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

### References

- 1. Wang, J.Q.; Cheng, G.; Tang, Z.P. The Distribution of Roughskin Sculpin (*Trachidermus fasciatus*) in Yalu River Basin, China. J. *Fudan Univ.* **2001**, *40*, 471–476. (In Chinese)
- Wang, J.Q. Advances in studies on the ecology and reproductive biology of *Trachidermus fasciatus* Heckel. *Acta Hydrobiol. Sin.* 1999, 23, 729–734. (In Chinese)
- 3. Yue, P.J.; Chen, Y.Y. China Red Data Book of Endangered Animal Species; Science Press: Beijing, China, 1988. (In Chinese)

- 4. Fireston, J.L.; Jasieniuk, M. Small population size limits reproduction in an invasive grass through both demography and genetics. *Oecologia* **2013**, *172*, 109–117. [CrossRef] [PubMed]
- Wang, J.Q.; Pan, L.D.; Liang, T.H.; Gan, H.C. Preliminary Study on Embryonic Development of the Roughskin Sculpin, *Trachider-mus fasciatus*. J. Fudan Univ. 2004, 2, 250–254. (In Chinese)
- Zhang, W.; Xu, J.R.; Xu, X.M.; Yin, X.J. Comparative experiments on different factory culture modes of *Trachidermus fasciatus*. *Fish. Mod.* 2017, 44, 6–9. (In Chinese)
- 7. Li, Y.L.; Xue, D.X.; Zhang, B.D.; Liu, J.X. Population genomic signatures of genetic structure and environmental selection in the catadromous roughskin sculpin *Trachidermus fasciatus*. *Genome Biol. Evol.* **2019**, *11*, 1751–1764. [CrossRef]
- 8. Ren, G.; Hu, J.; Bao, Z.; Gao, T. Development and characterization of fourteen microsatellite loci in a threatened catadromous fish *Trachidermus fasciatus. Conserv. Genet. Resour.* **2011**, *3*, 685–687. [CrossRef]
- 9. Bi, X.; Yang, Q.; Gao, T.; Li, C. The loss of genetic diversity during captive breeding of the endangered sculpin, *Trachidermus fasciatus*, based on ISSR markers: Implications for its conservation. *Chin. J. Oceanol. Limnol.* **2011**, *29*, 958–966. [CrossRef]
- 10. Zeng, J.; Liu, Z.Z.; Pan, L.D.; Tang, W.Q.; Wang, Q.; Geng, Y.H. Analysis of genetic diversity in wild populations of *Trachidermus fasciatus* by RAPD and the transformation of two SCAR markers. *Zool. Res.* **2012**, *33*, 203–210. (In Chinese) [CrossRef]
- 11. Liu, Z.Z.; Zeng, Z.; Pan, L.D.; Tang, S.J.; Yang, J.Q.; Tang, W.Q. Isolation and characterization of polymorphic microsatellite markers for the endangered roughskin sculpin (*Trachidermus fasciatus*). Conserv. Genet. Resour. 2012, 4, 837–840. [CrossRef]
- 12. Li, Y.L.; Xue, D.X.; Gao, T.X.; Liu, J.X. Genetic diversity and population structure of the roughskin sculpin (*Trachidermus fasciatus* Heckel) inferred from microsatellite analyses: Implications for its conservation and management. *Conserv. Genet.* **2016**, *17*, 921–930. [CrossRef]
- 13. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **2014**, *29*, 51–63. [CrossRef] [PubMed]
- 14. Whitehead, A. Comparative genomics in ecological physiology: Toward a more nuanced understanding of acclimation and adaptation. *J. Exp. Biol.* **2012**, *215*, 884–891. [CrossRef]
- 15. Nielsen, E.E.; Hemmer-Hansen, J.; Larsen, P.F.; Bekkevold, D. Population genomics of marine fishes: Identifying adaptive variation in space and time. *Mol. Ecol.* **2009**, *18*, 3128–3150. [CrossRef] [PubMed]
- 16. Xu, J.; Zhao, Z.; Zhang, X.; Zheng, X.; Li, J.; Jiang, Y.; Kuang, Y.; Zhang, Y.; Feng, J.; Li, C.; et al. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genom.* **2014**, *15*, 307. [CrossRef]
- Kjærner-Semb, E.; Ayllon, F.; Furmanek, T.; Wennevik, V.; Dahle, G.; Niemelä, E.; Ozerov, M.; Vähä, J.P.; Glover, K.A.; Rubin, C.J.; et al. Atlantic salmon populations reveal adaptive divergence of immune related genes—A duplicated genome under selection. BMC Genom. 2016, 17, 610. [CrossRef] [PubMed]
- Jian, J.; Yang, L.; Gan, X.; Wu, B.; Gao, L.; Zeng, H.; Wang, X.; Liang, Z.; Wang, Y.; Fang, L.; et al. Whole genome sequencing of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Hypophthalmichthys nobilis*) provide novel insights into their evolution and speciation. *Mol. Ecol. Resour.* 2021, 21, 912–923. [CrossRef]
- 19. Chen, D.; Zhang, Q.; Tang, W.; Huang, Z.; Zhang, J. The evolutionary origin and domestication history of goldfish (*Carassius auratus*). *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29775–29785. [CrossRef]
- 20. Xie, G.; Zhang, X.; Lv, F.; Sang, M.; Hu, H.; Wang, J.; Liu, D. Nanopore sequencing and Hi-C based de novo assembly of *Trachidermus fasciatus* genome. *Genes* **2021**, *12*, 692. [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. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, 25, 1754–1760. [CrossRef]
- Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079. [CrossRef] [PubMed]
- Narasimhan, V.; Danecek, P.; Scally, A.; Xue, Y.; Tyler-Smith, C.; Durbin, R. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 2016, 32, 1749–1751. [CrossRef] [PubMed]
- Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 2012, *6*, 80–92. [CrossRef] [PubMed]
- Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007, *81*, 559–575. [CrossRef]
- 27. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
- Zhang, C.; Dong, S.S.; Xu, J.Y.; He, W.M.; Yang, T.L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 2019, 35, 1786–1788. [CrossRef]
- 29. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009, 19, 1655–1664. [CrossRef]
- 30. Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **2020**, *36*, 2251–2252. [CrossRef]
- Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS J. Integr. Biol. 2012, 16, 284–287. [CrossRef]

- 32. Lorenzen, K.; Beveridge, M.C.M.; Mangel, M. Cultured fish: Integrative biology and management of domestication and interactions with wild fish. *Biol. Rev.* 2012, *87*, 639–660. [CrossRef] [PubMed]
- González, S.; Flick, G.J.; O'Keefe, S.F.; Duncan, S.E.; McLean, E.; Craig, S.R. Composition of farmed and wild yellow perch (*Perca flavescens*). J. Food Compos. Anal. 2006, 19, 720–726. [CrossRef]
- Arechavala-Lopez, P.; Fernandez-Jover, D.; Black, K.D.; Ladoukakis, E.; Bayle-Sempere, J.T.; Sanchez-Jerez, P.; Dempster, T. Differentiating the wild or farmed origin of Mediterranean fish: A review of tools for sea bream and sea bass. *Rev. Aquac.* 2013, 5, 137–157. [CrossRef]
- 35. Arechavala-Lopez, P.; Sanchez-Jerez, P.; Bayle-Sempere, J.T.; Sfakianakis, D.G.; Somarakis, S. Morphological differences between wild and farmed Mediterranean fish. *Hydrobiologia* **2012**, *679*, 217–231. [CrossRef]
- 36. López, M.E.; Cádiz, M.I.; Rondeau, E.B.; Koop, B.F.; Yáñez, J.M. Detection of selection signatures in farmed coho salmon (*Oncorhynchus kisutch*) using dense genome-wide information. *Sci. Rep.* **2021**, *11*, 9685. [CrossRef]
- 37. DeWoody, J.A.; Avise, J.C. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish. Biol.* **2000**, *56*, 461–473. [CrossRef]
- Tseng, M.C.; Chen, C.A.; Kao, H.W.; Tzeng, W.N.; Lee, S.C. Polymorphisms of GA/GT microsateilite loci from Anguilla japonica. Mar. Biotechnol. 2001, 3, 275–280. [CrossRef]
- Maes, G.E.; Pujolar, J.M.; Raeymaekers, J.A.M.; Dannewitz, J.; Volckaert, F.A.M. Microsatellite conservation and Bayesian individual assignment in four Anguilla species. *Mar. Ecol. Prog. Ser.* 2006, 319, 251–261. [CrossRef]
- 40. Vu, N.T.; Ha, T.T.T.; Thuy, V.T.B.; Trang, V.T.; Nguyen, N.H. Population genomic analyses of wild and farmed striped catfish *pangasianodon hypophthalmus* in the lower Mekong river. *J. Mar. Sci. Eng.* **2020**, *8*, 471. [CrossRef]
- Fagbémi, M.N.A.; Pigneur, L.M.; André, A.; Smitz, N.; Gennotte, V.; Michaux, J.R.; Mélard, C.; Lalèyè, P.A.; Rougeot, C. Genetic structure of wild and farmed Nile tilapia (*Oreochromis niloticus*) populations in Benin based on genome wide SNP technology. *Aquaculture* 2021, 535, 736432. [CrossRef]
- 42. Frankham, R. Genetics and extinction. Biol. Conserv. 2005, 126, 131–140. [CrossRef]
- Sremba, A.L.; Hancock-Hanser, B.; Branch, T.A.; LeDuc, R.L.; Baker, C.S. Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered antarctic blue whale (*Balaenoptera musculus intermedia*). *PLoS ONE* 2012, 7, e32579. [CrossRef] [PubMed]
- 44. Mbora, D.N.M.; McPeek, M.A. Endangered species in small habitat patches can possess high genetic diversity: The case of the Tana River red colobus and mangabey. *Conserv. Genet.* **2010**, *11*, 1725–1735. [CrossRef]
- 45. Ellis, J.R.; Pashley, C.H.; Burke, J.M.; McCauley, D.E. High genetic diversity in a rare and endangered sunflower as compared to a common congener. *Mol. Ecol.* 2006, *15*, 2345–2355. [CrossRef]
- 46. Pan, L.; Cai, F.; Ma, Z.; Jiang, X. Fisheries Science & Technology Information. Fish. Sci. Technol. Inf. 2010, 37, 211–214. (In Chinese)
- Yi, W.G.; Liang, P.; Liang, T.H.; Huang, Y.L. Roughskin Sculpin (*Trachidermus fasciatus*) in Yalu River Basin. *Hebei Fish.* 2006, 2, 53–54. (In Chinese)
- Oleksyk, T.K.; Smith, M.W.; O'Brien, S.J. Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. B Biol. Sci.* 2010, 365, 185–205. [CrossRef]
- 49. Bicskei, B.; Taggart, J.B.; Bron, J.E.; Glover, K.A. Transcriptomic comparison of communally reared wild, domesticated and hybrid *Atlantic salmon* fry under stress and control conditions. *BMC Genet.* **2020**, *21*, 57. [CrossRef]

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