Turk. J. Fish.& Aquat. Sci. 26(2), TRJFAS27297
https://doi.org/10.4194/TRJFAS27297

RESEARCH PAPER

Turkish Journal of FISHERIES and AQUATIC SCIENCES

Genetic Diversity and Structure of Largemouth Black Bass in Aquaculture Using Whole-genome SSR and SNP Markers

Kai Liu[1,*], Qinghui Meng[2,*], Wenrui Shi[2], Nan Xie[1]

[1]Hangzhou Academy of Agricultural Sciences, Hangzhou, China (310024).
[2]Zhejiang Fisheries Technical Extension Center, Hangzhou, China (310023).

How to Cite

Liu, K., Meng, Q., Shi, W., Xie, N. (2026). Genetic Diversity and Structure of Largemouth Black Bass in Aquaculture Using Whole-genome SSR and SNP Markers. *Turkish Journal of Fisheries and Aquatic Sciences*, *26(2)*, *TRJFAS27297.* https://doi.org/10.4194/TRJFAS27297

Article History
Received 19 November 2024
Accepted 04 July 2025
First Online 23 July 2025

Corresponding Author
E-mail: xqa9038@dingtalk.com
qinghui1234@126.com

Keywords
Genetic diversity
Ssr markers
Snp markers
Largemouth black bass

Abstract

Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs) offer distinct yet complementary insights into genetic diversity and population structure. This study compared whole-genome SSRs and SNPs to characterize Largemouth Black Bass (Micropterus salmoides) populations, a critical Chinese aquaculture species. We analyzed the selectively bred 'Zhejiang Black Bass No.1' F4 strain (ZL1f4) against two introduced cultured populations (Anhui: AL, Panzhihua: PL). High-throughput whole-genome resequencing yielded over one million high-quality SNP loci, and 28,250 (LobSTR) and 2,142 (SSRgenotyper) polymorphic SSRs, enabling robust comparative marker assessment. Results consistently demonstrated significantly reduced genetic diversity in selectively bred ZL1f4 (e.g., ZL1f4 average expected heterozygosity via LobSTR: 0.445 vs. AL: 0.536), reflecting artificial selection's genetic consequences. Both marker types revealed population differentiation, but SNPs consistently exhibited superior resolution (e.g., ZL1f4 vs. AL Fst: 0.1049 for SNPs vs. 0.0644 for LobSTR-SSR). Despite diversity differentiation, genetic structure analysis indicated retained shared ancestral components, suggesting selection influenced allele frequencies and heterozygosity, not fundamental genetic architecture. This research underscores SSR and SNP complementary strengths, advocating their integrated application for robust genetic characterization and informing sustainable aquaculture. Findings provide crucial data for effective genetic enhancement and long-term population management of Largemouth black bass, contributing to germplasm and aquatic biodiversity preservation.

Introduction

In molecular genetics, Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs) serve as critical genetic markers, each possessing distinct structural characteristics that confer unique advantages in genetic analysis. SSRs, also known as microsatellites, consist of short tandem repeats of one to six nucleotides. Their high polymorphism arises from variations in the number of these repeat units among individuals, driven by replication slippage, making them highly informative for detecting fine-scale genetic diversity and recent evolutionary events (Chen et al., 2024; Han et al., 2022). This inherent variability grants SSRs a high power of resolution for probing genetic diversity (Hodel et al., 2016). SNPs, in contrast, represent a difference at a single nucleotide position, making them the most prevalent form of genetic variation in a genome. Their immense abundance and typically uniform distribution across the genome provide a high resolution for assessing genetic linkage, identifying quantitative trait loci, and delineating population structure with high precision (Guo et al., 2022; Huang et al., 2020). The confluence of these markers' polymorphism and broad distribution, enhanced by molecular biology's technological advances, enables automated high-throughput analyses, thus bolstering the precision and efficiency of genetic studies. SSR and SNP markers are indispensable for identifying genes associated with crucial traits,

constructing genetic maps, evaluating genetic diversity, and elucidating population structures and evolutionary dynamics (Chen et al., 2022; Guo et al., 2022; Huang et al., 2020; Liu et al., 2023a).

Recent trends indicate a shift towards SNPs as the preferred genetic markers over SSRs due to their abundance and uniform genomic distribution, which facilitates the straightforward application of population genetic statistics (Tsykun et al., 2017). SNPs offer enhanced precision in estimating population diversity, superior analytical capabilities for cluster analysis, and the capacity to evaluate local adaptations (Zimmerman et al., 2020). Despite the increasing preference for SNPs, the combined use of both SSRs and SNPs in genetic research offers significant practical advantages due to their complementary roles. SSRs, with their high mutation rates and multi-allelic nature, are particularly powerful for detecting recent population bottlenecks, identifying closely related individuals, and tracking parentage (Chen et al., 2024; Han et al., 2022). In contrast, SNPs, with their high density and stability, are ideal for large-scale population genomic studies, genome-wide association studies, and fine-scale population structure analysis (Guo et al., 2022; Huang et al., 2020). Therefore, integrating both marker types can provide a more comprehensive and nuanced understanding of genetic diversity and population dynamics than either marker type alone. However, a significant knowledge gap persists in comprehensive comparative studies between SNPs and SSRs, particularly those utilizing whole-genome data. Existing research often compares a narrow set of SSR markers against comprehensive whole-genome SNP data (Szatmari et al., 2021; Zimmerman et al., 2020). For instance, earlier research by Zhang et al. (2014) utilized 18 SSRs for genetic clustering and distance analyses in *Megalobrama* fish, which diverged significantly from findings using whole-genome SSR and SNP data by Liu et al. (2023a) and Chen et al. (2022). Our previous studies further highlight the critical impact of the number of SSRs on research outcomes (Liu et al., 2023a), noting that a limited SSR panel may fail to capture total genomic variation, potentially skewing results. Consequently, this study aims to explicitly address this gap by conducting a comprehensive comparative analysis of whole-genome SSRs and SNPs to assess their relative effectiveness in evaluating genetic diversity and structure, thereby providing a more robust understanding of their utility in genetic studies.

Over recent decades, the Largemouth black bass (*Micropterus salmoides*), prized for its high-quality meat and economic value, has become central to China's freshwater aquaculture industry (Bai et al., 2008; Du et al., 2022). Native to North America's freshwater basins, this species was introduced to Guangdong Province in 1983 and has since proliferated across China. Despite its enhanced feeding and environmental adaptability post-domestication, the species contend with challenges such as limited germplasm resources, reduced genetic diversity, and inbreeding depression (Du et al., 2022). These challenges directly impact the sustainability and productivity of Largemouth black bass aquaculture. Our current study examines the F4 generation of 'Zhejiang Black Bass No.1,' a strain selectively bred for improved growth rates and regional adaptability. We aim to compare the genetic diversity and structure of this selected breeding population with that of the initial introductions, employing whole-genome SSR and SNP markers for a comprehensive genetic analysis. This empirical effort seeks to illuminate the impacts of artificial selection on the Largemouth black bass's genetic diversity and structure and to evaluate the distinct capacities of SSR and SNP markers in elucidating these aspects. The insights gained from this comparison are crucial for developing effective genetic enhancement strategies and informed population management practices, directly contributing to the mitigation of inbreeding depression and the long-term sustainability of aquaculture. Understanding the nuances between whole-genome SSR and SNP markers also enriches our comprehension of population genetics by offering more profound insights into genetic diversity and structure.

Materials and Methods

Experimental Samples

The Zhejiang Black Bass No.1 F4 generation (ZL1f4) was developed by the Zhejiang Province Fisheries Technology Extension Station in China through a selective breeding program targeting rapid growth. This lineage was derived from a naturally reproducing Largemouth black bass population harvested from the Bolongkeng Reservoir in Quzhou, Zhejiang Province, starting in 2017. The primary selection criterion for the breeding program was an enhanced growth rate. A selection pressure with an intensity of 0.2% was applied, focusing on growth rate and phenotypic traits.

Two cultured populations of the Northern subspecies of Largemouth black bass introduced from Florida, USA, have been established at the Xiba Base of the Zhejiang Province Fisheries Technology Extension Station. The Anhui population (AL) was introduced to China by Anhui Zhanglin Fisheries Co., Ltd., and the Panzhihua population (PL) was introduced by Panzhihua Laibei Fisheries Co., Ltd. in Sichuan, China. Since their introduction, both populations have developed into naturally reproducing inbred lines.

For genetic analysis, ten individuals from each population were randomly selected. These individuals were all two years old and weighed between 500 and 800 grams. Pectoral fins were collected and preserved in anhydrous ethanol. All procedures involving fish were approved by the administrative committees of the Institute of Fishery Science at the Hangzhou Academy of Agricultural Sciences and the Zhejiang Fisheries Technical Extension Center. Genomic DNA was then

extracted using the phenol-chloroform method (Green, Sambrook, 2018). While a sample size of ten individuals per population might be considered relatively small for comprehensive population-level genetic diversity estimation in whole-genome sequencing studies, this choice was made to balance the high cost and computational demands of whole-genome resequencing with the need for high-quality, representative data from each distinct population. The subsequent stringent data filtering steps (detailed in Sections 2.3 and 2.4) were critical to ensure the reliability of the genetic markers used, compensating for the sample size limitation by focusing on high-confidence genotypes across all samples.

Library Construction, Sequencing, and Data Quality Control

Library construction, sequencing, and data quality control were rigorously performed in adherence to the Illumina protocol for whole-genome sequencing designed for Next Generation Sequencing library preparation (Quail et al., 2008). Quantified genomic DNA was fragmented using a Covaris M220 sonicator, followed by several purification steps to prepare the sequencing library. This preparation involved end-repair of DNA fragments, adenylation of 3' ends, and ligation of sequencing adapters, culminating in bridge PCR amplification. Per the manufacturer's instructions, the libraries were constructed using the Illumina TruSeq sample preparation kit (Illumina, USA). Sequencing on an Illumina NextSeq 550 platform produced 300 base pair paired-end reads. The sequencing effort was undertaken by Wuhan Frasergen Bioinformatics Co., Ltd, China. Quality control for the generated raw reads was executed using Fastp v0.23.1 (Chen, 2023) with parameters set to -c -D, ensuring the acquisition of high-quality, clean reads.

Whole-genome SNP Genotyping

Using the Largemouth black bass genome GCA_022435785.1 (He et al., 2022) as the reference genome, clean reads were aligned to the reference using bwa-meme v1.0.5 (Jung, Han, 2022), producing SAM format files. These SAM files were converted to BAM format using samtools v1.6 (Danecek et al., 2021), and duplicate reads were removed with sambamba v1.0 (Tarasov et al., 2015). Insertions and deletions (indels) realignment was performed using Picard v2.27.5 (Broad, 2019) and GATK v3.8.1 (DePristo et al., 2011) for local realignment of reads. Samtools v1.6 was used again to filter out improperly aligned reads (-F 0x904 -q 30). VCF files were generated using bcftools v1.8 (Danecek et al., 2021) mpileup and call commands, and the resulting VCF files were filtered with the filter command to remove variants with allele bias >90%, quality <20, and depth <10. These filtering steps are crucial to ensure the reliability and accuracy of SNP calls by removing low-quality variants that could arise from sequencing errors or mapping artifacts. Specifically, removing variants with allele bias >90% addresses strand bias, quality <20 removes low-confidence calls, and depth <10 ensures sufficient read coverage for robust genotype determination. Further filtering of the VCF files was conducted using vcftools v0.1.16 (Danecek et al., 2011) to exclude variants with a minor allele frequency (MAF) less than 0.05, quality score below 10, depth below 10, all missing genotypes, and all indel types. The MAF<0.05 filter removes rare variants that may not be informative for population-level analyses and could be artifacts. Filtering for quality score below 10 and depth below 10 further enhances the stringency of SNP selection, ensuring only high-confidence, well-supported loci are retained. Removing all missing genotypes ensures that only markers present in all samples are considered, which is vital for comparative analyses across populations. The final SNP genotyping results were saved in VCF format.

Vcftools v0.1.16 was employed to calculate several genetic metrics for each population, including the number of SNP loci, observed heterozygosity (*Ho*), expected heterozygosity (*He*), nucleotide diversity (π), and minor allele frequency (MAF). We were using adegenet (Jombart, Ahmed, 2011), vcfR (Knaus, Grünwald, 2017), and StAMPP (Pembleton et al., 2013), with the stamppFst function specifically used to calculate Weir and Cockerham's Fst value (wcFst) (Weir, Cockerham, 1984). Additionally, the inbreeding coefficient (Fis) was determined using the --het option in Plink v1.9 (Purcell et al., 2007).

Whole-genome SSR Genotyping

To conduct whole-genome SSR genotyping, we employed two distinct tools: LobSTR v4.0.6 (Gymrek et al., 2012) and SSRgenotyper (Lewis et al., 2020). The rationale for utilizing both tools was to perform a comprehensive comparative assessment of their performance in identifying and genotyping whole-genome SSRs from resequencing data, and subsequently, to compare the genetic insights derived from these SSR datasets with those obtained from SNP markers. This comparative approach is central to addressing the research gap identified in the Introduction regarding the effectiveness of different marker types and genotyping methodologies. LobSTR focuses on utilizing signal processing techniques to identify and characterize SSR sequences, enhancing the accuracy and efficiency of SSR analysis by specifically addressing noise generated during the PCR amplification process. SSRgenotyper, in contrast, emphasizes direct identification and genotyping of SSR loci from resequencing data by aligning sequencing reads to SSR reference sequences and determining genotypes based on the proportion of reads that support each allele.

To facilitate whole-genome SSR genotyping of the Largemouth black bass genome GCA_022435785.1 (He

et al., 2022), two distinct methodologies were employed. Initially, Tandem Repeats Finder (TRF) v4.09 (Benson, 1999) was utilized to identify SSRs. The parameters were set as follows: a matching weight of 2, a mismatch penalty of 7, an indel penalty of 7, a match probability of 80, an indel probability of 10, a minimum alignment score to report of 50, and a maximum period size to report of 500. Additional settings included the use of flanking sequences (-f), the creation of a data file (-d), and a masked sequence file (-m). Data from the TRF output were processed using a Python script, GetSTRInfo.py, from LobSTR v4.0.6, facilitating the generation of a BED file defining SSR regions compatible with LobSTR analysis. We used the allelotype function of LobSTR v4.0.6 for SSR genotyping, analyzing BAM files derived from the SNP genotyping phase that had undergone indel realignment and filtering of improperly aligned reads. The SSR genotyping results were stored in VCF format. The VCF file was further processed using the lobSTR_filter_vcf.py script to exclude loci based on specific criteria: maximum reference length of a locus at 80, minimum mean log score cutoff at 0.8, minimum mean coverage at 5, and a minimum call rate of 1.0. These stringent filtering criteria for LobSTR were applied to ensure high-confidence SSR genotype calls, minimizing false positives and ensuring that only robustly genotyped loci are included in downstream analyses. A call rate of 1.0, for instance, means that only SSRs successfully genotyped in all individuals were retained, which is critical for direct comparisons across populations. The SSR genotyping data were saved in VCF format and subsequently converted to GENEPOP format using a custom script, available at https://github.com/zergger/SSRgenotyper.

Subsequently, SSRs from the same genome were mined using MISA software (Thiel et al., 2003), focusing on di-nucleotide repeats appearing six or more times and tri- to hexanucleotide repeats occurring at least four times. Clean reads were aligned to these SSR-containing sequences using bwa-meme v1.0.5 (Jung, Han, 2022), and SAM format files were converted to BAM format files using samtools v1.6 (Danecek et al., 2021) with quality control (-q 30). Duplicate reads were removed using sambamba v1.0 (Tarasov et al., 2015) for SSRgenotyper processing. Whole-genome SSR genotyping was conducted using SSRgenotyper, with parameters set to Q=30, S=1, M=0.1, F=0.01, B=50, and m=5. These parameters ensure that only high-quality reads (Q=30) are considered for genotyping, and that heterozygous calls are supported by a minimum allele percentage (M=0.1) and read count (S=1). The maximum missing data threshold (F=0.01) ensures a high call rate for retained loci, while flanking sequence parameters (B=50, m=5) help in accurate alignment and genotyping of SSR regions. Additionally, we modified parts of the SSRgenotyper code to read results from MISA directly. The modified code is available at https://github.com/zergger/SSRgenotyper. The resulting SSR genotyping data was saved in GENEPOP format.

Allelic diversity metrics, such as the number of alleles (*N*a), allelic richness (*A*r), observed heterozygosity (*H*o) and expected heterozygosity (*H*e), were calculated for each population using the R package adegenet, strataG (Archer et al., 2017). The pairwiseTest function was employed to compute Fis between populations. Using adegenet, dartR (Mijangos et al., 2022), and StAMPP, with the stamppFst function specifically used to calculate wcFst.

Genetic Diversity Analysis

Genetic distance analyses among populations employ Nei's genetic distance (Nei, 1972). For SSR data, these calculations are facilitated by the R packages adegenet, dartR, and StAMPP, utilizing the stamppNeisD function to compute genetic distances. For SNP data, the analysis uses adegenet, vcfR, and StAMPP, with the stamppNeisD function specifically used to calculate genetic distances. Crucially, all subsequent genetic diversity and structure analyses were performed independently for the datasets generated by LobSTR, SSRgenotyper, and SNP genotyping. This approach allowed for a direct and unbiased comparison of the resolution and insights provided by each marker type and genotyping methodology across various population genetic analyses. We also assess the correlation between genetic distances measured using LobSTR, SSRgenotyper, and SNP markers. The Mantel Test was performed using the mantel.test function from the Python package Mantel, employing the Pearson method with 10,000 permutations (Mantel, 1967). The Mantel Test is a statistical tool used to assess the correlation between two distance matrices, in this context, to determine if genetic distances calculated using different marker types (SSR vs. SNP) are congruent. A significant correlation indicates that different markers capture similar patterns of genetic differentiation among individuals. Additionally, linear regression analysis was conducted using the Ordinary Least Squares (OLS) method (Rohlf, Sokal, 1995) from the Python package statsmodels. OLS regression was employed to quantitatively evaluate the predictive relationship between genetic distances derived from SSR markers (LobSTR and SSRgenotyper) and those from SNP markers. This allows us to determine how well SSR-based distances can explain the variation observed in SNP-based distances, providing insights into the comparative utility of these marker types for genetic distance estimation.

Principal Coordinates Analysis (PCoA) is conducted based on genetic distance matrices among individuals. The Python package scikit-bio, specifically its pcoa function, clusters these genetic distance matrices into three principal coordinate axis components. Subsequently, visualization is achieved through the use of the matplotlib plotting function.

Utilizing the aboot function from the R package poppr (Kamvar et al., 2015), we constructed

phylogenetic trees using the Neighbor-Joining (NJ) method based on genetic distances among individuals. The trees' confidence levels were validated through Bootstrap analysis with 1,000 resamplings. For visual representation, the R package ggtree (Yu et al., 2016) was employed. Additionally, the poppr.msn function from poppr was used to construct phylogenetic networks through the Minimum Spanning Network method.

Prior to analyzing the genetic structure of populations, we employed the poppr.amova function from the R package poppr, which utilizes the pegas method (Paradis, 2010) for analyzing molecular variance based on genetic distances among individuals. This analysis included 1,000 random permutations to test for significant genetic differentiation between populations robustly. In analyzing SSR data, we utilized Structure v2.3.4 (Hubisz et al., 2009) to assess genetic structure, testing K values from one to five, each with five iterations. The most suitable K was identified using KFinder v1.0 (Wang, 2019). To integrate results from multiple runs, Clumpp v1.1.2 (Jakobsson, Rosenberg, 2007) was used to generate a Q matrix, which was visualized with distruct v1.1 (Rosenberg, 2003). For SNP analysis, loci devoid of solid linkage disequilibrium were selected via Plink v1.9, applying the settings --indep-pairwise 100 10 0.5. Genetic structure was then determined using admixture v1.3 (Alexander et al., 2009), exploring K values from one to five with variable seeds and repeated five times each. Optimal K values were ascertained using a cross-validation approach (Alexander, Lange, 2011). Finally, genetic structure plots were created using the pong Python package (Behr et al., 2016).

Results

Genotyping Analysis

A total of 30 individuals underwent whole-genome resequencing, achieving a Q30 quality score of over 94% and an average sequencing depth exceeding 9x. For details on sequencing data quality control, please refer to Supplementary Table S1. Results concerning sample sequencing depth and coverage are available in Supplementary Table S2. Moreover, during the genotyping process, we did not impute missing genotypes. Instead, we retained only those markers that were genotyped across all samples. For instance, for genotyping with LobSTR, we required a call rate of 1.0; for SSRgenotyper, we set a maximum missing data threshold of 0.01 (F=0.01). For SNPs, we filtered out all missing genotypes.

Using LobSTR, the ZL1f4 population, and AL, PL populations were genotyped for SSR markers through sequencing methods, identifying 28,250 polymorphic SSR loci. Of these, 380 were mono-nucleotide repeats, 22,460 were di-nucleotide repeats, and the remaining comprised 2,354 tri-nucleotide, 2,406 tetra-nucleotide, 787 penta-nucleotide, and 635 hexa-nucleotide repeats. It is important to note that while mono-nucleotide repeats can be susceptible to sequencing errors, the stringent filtering criteria applied (e.g., minimum mean log score cutoff at 0.8, minimum mean coverage at 5, and a minimum call rate of 1.0) were designed to minimize such artifacts and ensure high-confidence calls for all SSR types. The density distribution of SSRs across the chromosomes is shown in Supplementary Figure S1. The *N*a detected across the three populations has an average of 5.696. The ZL1f4 population had a relatively lower *H*e of 0.445, whereas the AL population had a higher *H*e of 0.536. The distribution of other metrics followed a similar pattern to that of *H*e, with AL>PL>ZL1f4 (see Supplementary Table S3). These observed differences in genetic diversity, with the ZL1f4 population showing lower heterozygosity, are significant as they indicate a potential impact of the selective breeding program on the genetic makeup of this cultured strain. Reduced heterozygosity can be a precursor to inbreeding depression, which negatively affects fitness traits in aquaculture populations.

Using SSRgenotyper, the ZL1f4 population, and AL, PL populations were genotyped for SSR markers through sequencing methods, identifying 2,142 polymorphic SSR loci. Of these, 1,837 were di-nucleotide repeats, and the remaining comprised 183 tri-nucleotide, 113 tetra-nucleotide, eight penta-nucleotide, and one hexa-nucleotide repeats. The density distribution of SSRs across the chromosomes is shown in Supplementary Figure S2. The *N*a detected across the three populations ranged from two to six, with an average *N*a of 2.346. Regarding *H*e, the ZL1f4 population exhibited a relatively lower *H*e at 0.122, while the AL population displayed a higher *H*e at 0.137. The distribution of *H*e across populations followed a distinct pattern, with AL>PL>ZL1f4. In contrast, the other three metrics—*N*a, *A*r, and *H*o—showed slight variations in their distributions, presenting as AL>ZL1f4>PL (see Supplementary Table S4). The type and number of polymorphic repeats identified by LobSTR (28,250 loci) and SSRgenotyper (2,142 loci) are indeed grossly different. This significant discrepancy underscores the distinct underlying algorithms and filtering approaches of these two software tools in identifying and genotyping SSRs from the same whole-genome resequencing data. While both are designed for SSR analysis, their methodologies lead to variations in the detected polymorphism spectrum.

Using the publicly available Largemouth black bass genome on NCBI as a reference, genome-wide mining revealed an average of 1,119,539.033 SNP loci per individual (see Supplementary Table S5). The density distribution of SNPs across the chromosomes is shown in Supplementary Figure S3. The distribution of SNP genotyping results for MAF, π, *H*o, and *H*e followed a similar pattern to the *H*e distribution observed in SSR genotyping results, with AL>PL>ZL1f4. The high number of identified SNP loci underscores their extensive

genomic coverage and potential for high-resolution genetic analysis, providing a robust baseline for comparative assessments.

Based on *Ho* and *He*, we compared genotyping results from LobSTR, SSRgenotyper, and SNPs. The significance of the differences was assessed using a permutation test with 10,000 iterations. As illustrated in Figure 1, the genotyping results from LobSTR closely resemble those from SNPs, while those from SSRgenotyper exhibit substantial deviations. This analysis highlights the comparative accuracy and deviations in genotyping methods applied to the same genomic data. The closer alignment of LobSTR results with SNP data suggests that LobSTR may provide a more accurate representation of overall genomic diversity compared to SSRgenotyper, which appears to underestimate heterozygosity. This discrepancy is likely due to methodological differences in how each tool identifies and genotypes SSRs from whole-genome resequencing data, potentially related to read alignment stringency or repeat motif identification algorithms.

Additionally, we assessed genetic differentiation and inbreeding coefficient among populations based on genotyping results from LobSTR, SSRgenotyper, and SNPs, using wcFst and Fis (see Supplementary Table S6 and Figure 2). Results depicted in Figure 2 indicate that the wcFst values calculated from SNP data are higher than those derived from LobSTR and SSRgenotyper. Specifically, the wcFst values between the ZL1f4 and AL populations, based on SNP, LobSTR, and SSRgenotyper genotyping, were 0.1049, 0.0644, and 0.0463, respectively. According to Wright (1978), an Fst value of less than 0.15 suggests that most population differentiation is moderate, reflecting relatively minor genetic discrepancies among populations. The higher Fst values obtained from SNP data suggest that SNPs are more effective at detecting subtle population differentiation, likely due to their higher density and more uniform distribution across the genome compared to SSRs. This implies that while all markers indicate moderate differentiation, SNPs provide a more sensitive measure of genetic divergence between the selected breeding population (ZL1f4) and the introduced cultured populations (AL, PL). Additionally, Figure 2 shows that Fis values based on LobSTR and SNPs are approximately zero (Figure 2A and 2C), suggesting minimal disparity between observed and expected homozygosity, indicating negligible inbreeding or outbreeding. The values based on LobSTR are slightly negative, while those derived from SNPs vary, being both negative and positive across different populations. In contrast, Fis values estimated from SSRgenotyper significantly deviate from zero and are more significant than zero (Figure 2B), indicating the presence of inbreeding within these populations. The discrepancy in Fis values, particularly the significant positive values from SSRgenotyper, highlights a potential overestimation of inbreeding or a bias in genotype calling by this tool, which warrants careful consideration when interpreting results related to population genetic health.

Genetic Distance Analysis

We evaluated the variability and correlation in estimates of standard Nei's genetic distances based on genotyping results from LobSTR, SSRgenotyper, and SNPs (Supplementary Table S6 and Figure 3). According to the data in Supplementary Table S6, the genetic distances estimated from SNP genotyping results were the largest, followed by those estimated from LobSTR genotyping. Estimates from SSRgenotyper yielded the smallest genetic distances. Specifically, the genetic

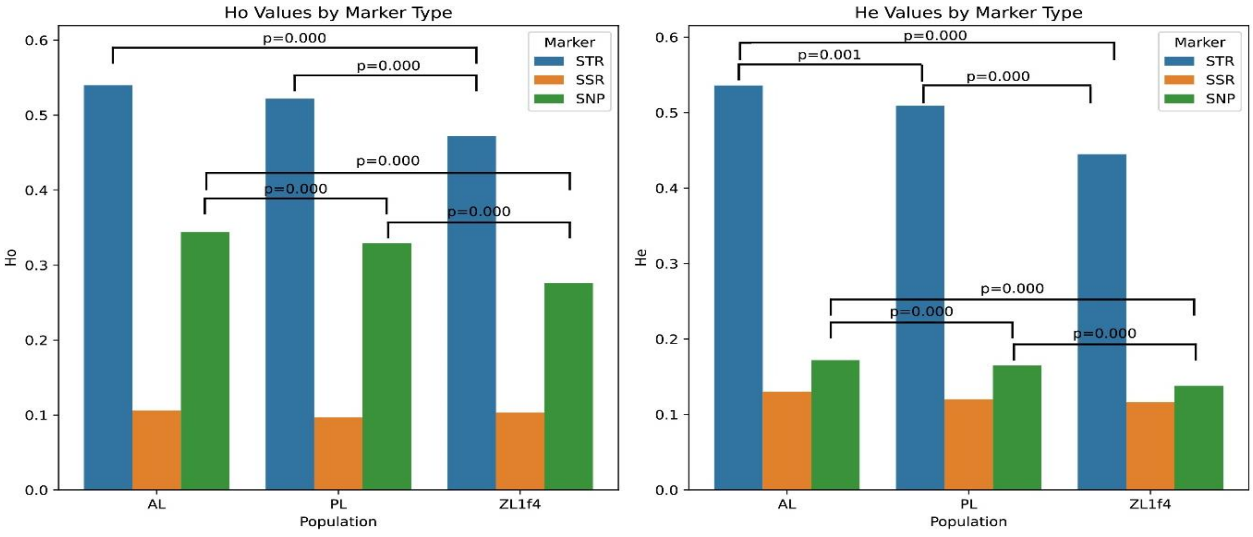

Figure 1. Comparison of Observed Heterozygosity (Ho) and Expected Heterozygosity (He) across Largemouth Black Bass Populations by Genotyping Method. (A) Observed Heterozygosity (Ho) values for Anhui (AL), Panzhihua (PL), and Zhejiang Black Bass No.1 F4 (ZL1f4) populations. (B) Expected Heterozygosity (He) values for AL, PL, and ZL1f4 populations. In both panels, values are presented for markers genotyped by LobSTR (blue bars), SSRgenotyper (orange bars), and SNP (green bars). P-values indicate significant differences between marker types within each population.

distances between the ZL1f4 and AL populations, based on SNP, LobSTR, and SSRgenotyper genotyping, were 0.0685, 0.0241, and 0.0138, respectively. These findings suggest that SNP markers provide a broader range of genetic distance estimates, potentially reflecting a more comprehensive capture of genetic variation across the genome. The smaller distances from SSRgenotyper further support its lower sensitivity in detecting overall genetic diversity compared to LobSTR and SNPs.

Based on genetic distances between individuals, the Mantel Test analysis revealed a significant correlation among distance matrices derived from LobSTR, SSRgenotyper, and SNPs (Figure 3). Specifically, the correlation coefficient from SSRgenotyper was 0.771 (P-value=0, Figure 3B), higher than the coefficient of 0.546 from LobSTR (P-value=0, Figure 3A). Utilizing the genetic distances from LobSTR or SSRgenotyper as independent variables and those from SNP markers as dependent variables, the OLS regression analysis based on SSRgenotyper demonstrated that the linear regression model accounted for 59.5% of the variability in the dependent variable (R-squared value=0.595, P-value=0, Figure 3B). This predictive capability was higher than that based on LobSTR, where the R-square d value was 0.298 (P-value=0, Figure 3A). The strong correlation observed, particularly with SSRgenotyper, indicates that despite differences in absolute genetic distance values, all three marker types capture similar underlying patterns of genetic relatedness among individuals. The higher predictive capability of SSRgenotyper in OLS regression, despite its lower diversity estimates, suggests it might be capturing a specific, highly correlated subset of genomic variation that aligns well with SNP-based distances.

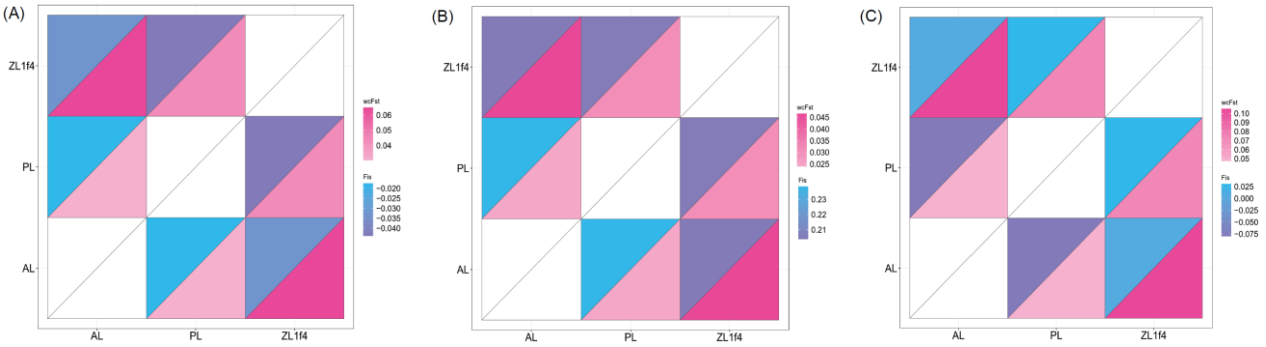

Figure 2. Comparison of Genetic Differentiation (wcFst) and Inbreeding Coefficient (Fis) Among Largemouth Black Bass Populations, Assessed by Different Genotyping Methods. Each matrix illustrates wcFst values in the lower triangle (light pink to red scale) and Fis values in the upper triangle (light blue to dark blue scale). (A) Parameters derived from LobSTR markers. (B) Parameters derived from SSRgenotyper markers. (C) Parameters derived from SNP markers. All panels depict genetic relationships among the selectively bred ZL1f4, and introduced Anhui (AL) and Panzhihua (PL) cultured populations.

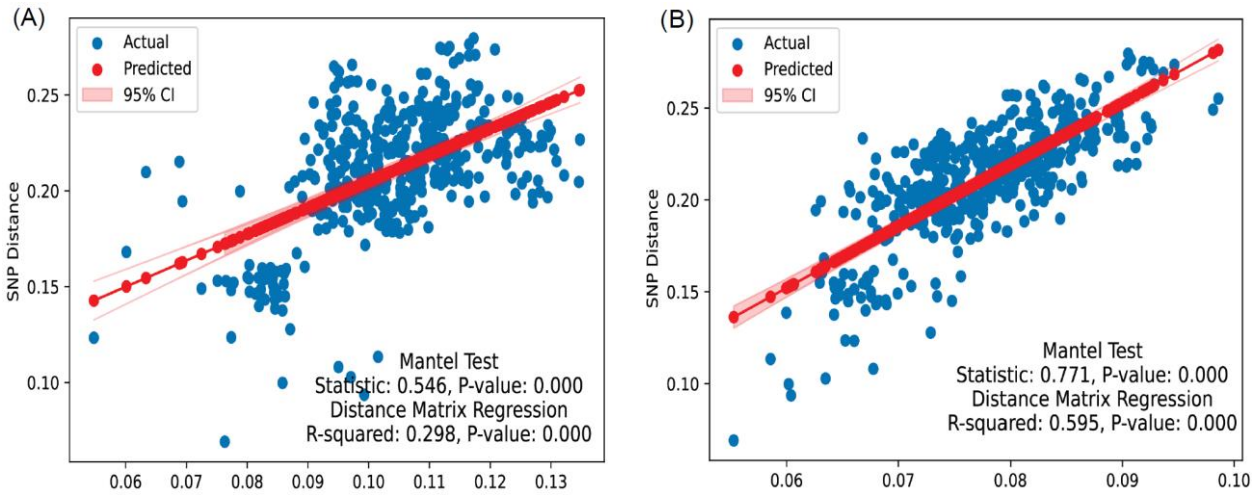

Figure 3. Correlation between Standard Nei's Genetic Distances Derived from SSR (LobSTR and SSRgenotyper) and SNP Markers. (A) Correlation between LobSTR-derived SSR distances and SNP distances. (B) Correlation between SSRgenotyper-derived SSR distances and SNP distances. Both panels display actual data points (blue), predicted regression line (red), and 95% confidence interval (pink shaded area), along with Mantel Test statistics and OLS regression R-squared values.

Genetic Diversity Analysis

Principal Coordinates Analysis

PCoA analysis using SSR markers derived from LobSTR delineates the ZL1f4 population from the introduced AL and PL populations, as depicted in Figure 4A. Nonetheless, some overlap between ZL1f4 and PL populations persists, highlighting the limited genetic resolution achieved with 28,250 SSR markers. This resolution is inadequate for the complete differentiation of these populations. Notably, the ZL1f4 population samples are more tightly clustered and exhibit less variance in pairwise distances, suggesting the impacts of selective breeding. In contrast, analysis using SSR markers from SSRgenotyper (Figure 4B) reveals substantial overlap between the ZL1f4 and PL populations, starkly contrasting the findings in Figure 4A. This discrepancy underscores the variable effectiveness of different SSR genotyping tools in distinguishing genetic differences among populations. The inability of SSRgenotyper to clearly differentiate populations, in contrast to LobSTR, reinforces the idea that SSRgenotyper may be less sensitive or accurate in capturing fine-scale genetic distinctions, potentially due to its methodology or the number of polymorphic loci it identifies.

Additionally, a PCoA analysis utilizing SNP markers (Figure 4C) reveals a clear distinction between the ZL1f4 and PL populations. Using a substantially more extensive set of SNP markers than SSR markers significantly enhances the genetic resolution, allowing for complete differentiation of these populations. Consistent with traits associated with selective breeding, the ZL1f4 samples are more tightly grouped, demonstrating minimal variance in inter-sample distances. This result highlights the robust capability of SNP markers in resolving genetic distinctions within populations. The superior resolution provided by SNP markers in PCoA is a key finding, demonstrating their efficacy in precisely delineating genetic relationships and the impact of artificial selection on population structure, which is crucial for effective breeding management.

Phylogenetic Analysis

The phylogenetic trees generated from SSR and SNP markers, as shown in Figure 5, reveal that the ZL1f4 population and two other populations, AL and PL, typically form three separate clades, although some samples exhibit intermingling branches. Notably, the branch lengths in the phylogenetic tree derived from SNP markers (Figure 5C) are more prolonged than those observed in the SSR-based tree generated by SSRgenotyper (Figure 5B). However, these lengths are comparable to those in the tree constructed using SSR markers identified with LobSTR (Figure 5A). Despite these similarities, the trees in Figure 5 display considerable divergence. While Figures 5A and 5C show a certain resemblance, they highlight the distinct tree structures that arise from using different genetic markers. The consistent formation of distinct clades, despite some intermingling, indicates underlying genetic differentiation between the selected and introduced populations. The longer branch lengths in the SNP-based tree further suggest that SNPs capture a greater extent of genetic divergence, providing a more refined view of evolutionary relationships.

The phylogenetic networks generated using SSR and SNP markers, as illustrated in Figure 6, reveal distinct outcomes compared to the corresponding phylogenetic trees. Networks constructed with SSR markers, employing SSRgenotyper (Figure 6B), show the ZL1f4 population intersecting with those of two introduced cultured populations, thus forming a tri-branched structure. In contrast, the networks constructed from both SSR (using LobSTR) and SNP markers (Figures 6A and 6C) form a predominant, expansive network that includes the ZL1f4 population and the two introduced populations, with only a few

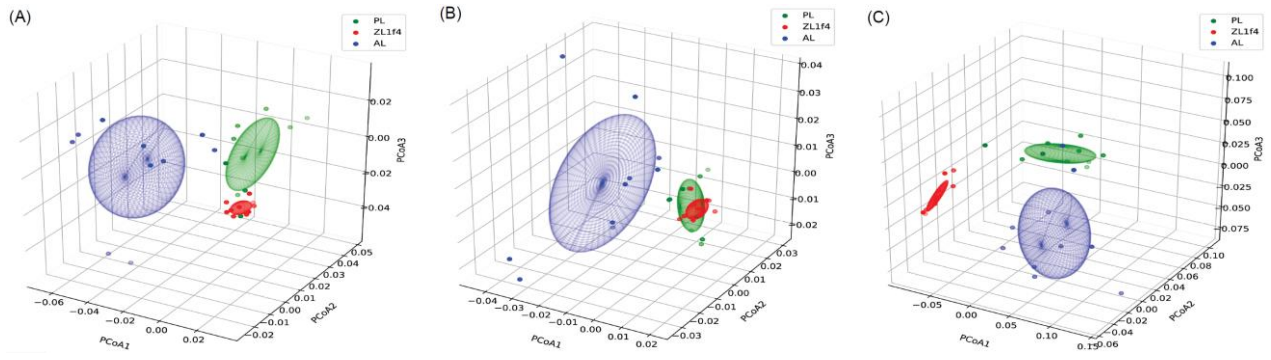

Figure 4. Principal Coordinates Analysis (PCoA) of Largemouth Black Bass Populations Based on Genetic Distances Derived from Different Marker Types. (A) PCoA results using LobSTR-derived SSR markers. (B) PCoA results using SSRgenotyper-derived SSR markers. (C) PCoA results using SNP markers. Each panel displays the genetic relationships among the selectively bred 'Zhejiang Black Bass No.1' F4 population (ZL1f4, red), and the introduced Anhui (AL, blue) and Panzhihua (PL, green) cultured populations. Ellipsoids represent the standard deviation ellipsoids for each population cluster, illustrating their genetic spread in the PCoA space.

samples diverging into minor branches. While Figures 6A and 6C demonstrate similarities, significant variations are evident, particularly the central positioning of the ZL1f4 population in the network constructed with LobSTR, highlighting notable discrepancies in the data derived from SSR and SNP markers. These network analyses provide complementary insights to the phylogenetic trees, confirming the genetic relatedness but also revealing the complex interconnections within and between populations. The differences observed between SSRgenotyper and the other two marker types again underscore the importance of marker choice and genotyping methodology in shaping the perceived genetic landscape.

Genetic Structure Analysis

Analysis of molecular variance based on genetic distances among individuals revealed profound differences between the ZL1f4 population and the AL and PL populations, as indicated by P-values of 0.0 for both SSR and SNP markers. SSR markers, analyzed using either LobSTR or SSRgenotyper, attributed 18.17% and 13.97% of the total genetic variance to inter-population differences, with Phi-statistics of 0.1665 and 0.1066, respectively. In contrast, SNP marker analysis showed that 25.78% of the variance stemmed from differences among populations, as evidenced by a Phi-statistic of 0.2695. For further details on these analyses, please consult Supplementary Tables S7, S8, and S9. These significant Phi-statistics consistently demonstrate genetic differentiation between the selected ZL1f4 population and the introduced populations, with SNPs revealing a higher proportion of variance attributed to inter-population differences, further supporting their higher resolution in detecting population structure.

Subsequently, genetic structure analyses were conducted using both SSR and SNP markers. Figure 7 illustrates the genetic structure plots for K=1, 2, and 3. Employing the methods of Pritchard et al. (2000), Evanno et al. (2005), and Wang (2019), the optimal K value was determined to be two when using LobSTR

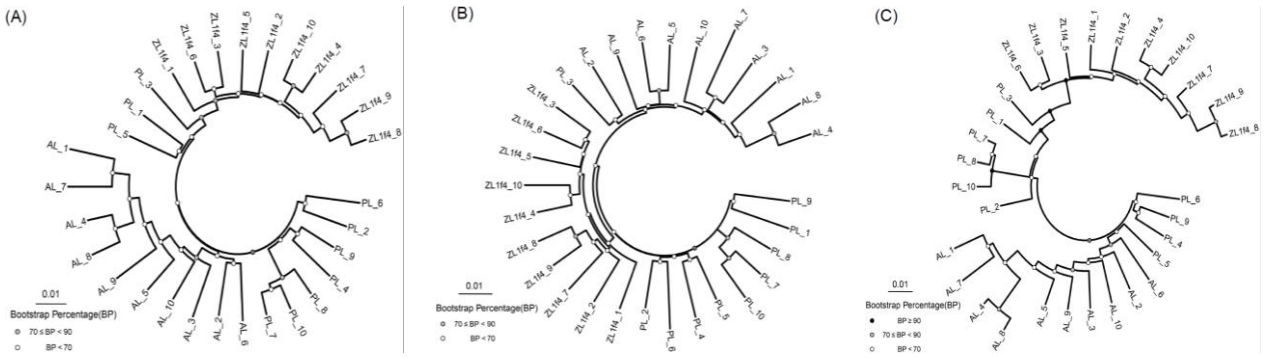

Figure 5. Phylogenetic Trees Illustrating Genetic Relationships Among Largemouth Black Bass Populations, Constructed Using Different Marker Types. (A) Neighbor-Joining phylogenetic tree derived from LobSTR-genotyped SSR markers. (B) Neighbor-Joining phylogenetic tree derived from SSRgenotyper-genotyped SSR markers. (C) Neighbor-Joining phylogenetic tree derived from SNP markers. All trees depict the genetic relationships among the selectively bred 'Zhejiang Black Bass No.1' F4 population (ZL1f4) and the introduced Anhui (AL) and Panzhihua (PL) cultured populations. Bootstrap percentages (BP) are indicated by circle color: white circles for BP<70, grey circles for 70≤BP<90, and black circles for BP≥90. The scale bar represents the genetic distance.

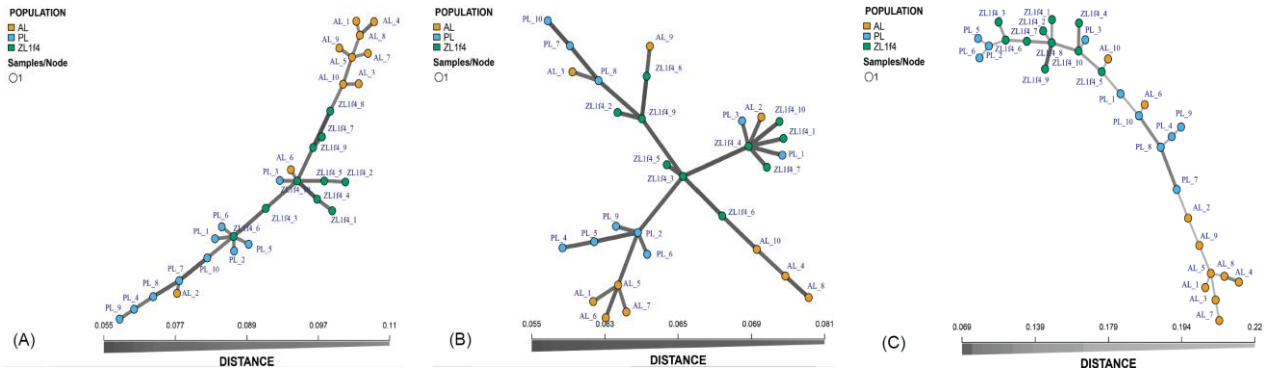

Figure 6. Phylogenetic Networks Illustrating Genetic Relationships Among Largemouth Black Bass Populations, Constructed Using Different Marker Types. All networks were constructed using the Minimum Spanning Network method. (A) Network based on LobSTR-genotyped SSR markers. (B) Network based on SSRgenotyper-genotyped SSR markers. (C) Network based on SNP markers. In all panels, nodes represent individual samples, colored according to population: AL (orange), PL (blue), and ZL1f4 (green). The length of the branches indicates genetic distance. The scale bar represents the genetic distance.

(Figure 7A). In contrast, analyses with SSRgenotyper confirmed a K value of two according to methods established by Pritchard et al. (2000), and Evanno et al. (2005), but the parsimony approach of Wang (2019) indicated that the optimal K value was one (Figure 7B). Furthermore, cross-validation results from SNP marker analyses revealed the lowest CV error, 0.5834±0.0004, across five runs for K=1, suggesting that this was the optimal K value for SNP markers (Figure 7C). However, we observed that the CV error for K=2, 0.6003±0.0024, was slightly higher than for K=1. The differing optimal K values derived from various methods (K=2 for LobSTR, K=1 or K=2 for SSRgenotyper depending on method, and K=1 for SNPs based on cross-validation) suggest that each marker type and analysis algorithm may capture different aspects of population substructure. While K=1 was indicated as optimal for SNPs, the relatively close CV error for K=2 (0.6003±0.0024) suggests that a subtle underlying structure might still exist, or that the methods are sensitive to different scales of genetic variation. If K=2 is considered the optimal value, then the genetic structure of the ZL1f4 population differs significantly from that of the AL and PL populations, indicating a more homogeneous genetic structure within the ZL1f4 population. In Figure 7, regardless of whether K equals one or two, the ZL1f4 population and the AL and PL populations appear to share the same ancestral component for LobSTR and SSRgenotyper. This observation aligns with the origin of the species for these three populations, all of which are northern subspecies of Largemouth black bass. Notably, artificial selection has not significantly modified the genetic structure of ZL1f4. The overall consistency in ancestral components across populations, despite some differentiation, indicates that the selective breeding program for ZL1f4 has primarily influenced genetic diversity rather than fundamentally altering the broader genetic structure, which is a positive outcome for germplasm conservation.

Discussion

The primary objectives of this study were to (1) assess the capabilities of whole-genome SSR and SNP markers in elucidating the genetic diversity of the selectively bred 'Zhejiang Black Bass No.1' F4 generation (ZL1f4) relative to two introduced cultured populations (AL and PL) and (2) evaluate the genetic consequences of the selective breeding program on genetic diversity alterations within the ZL1f4 population. Our

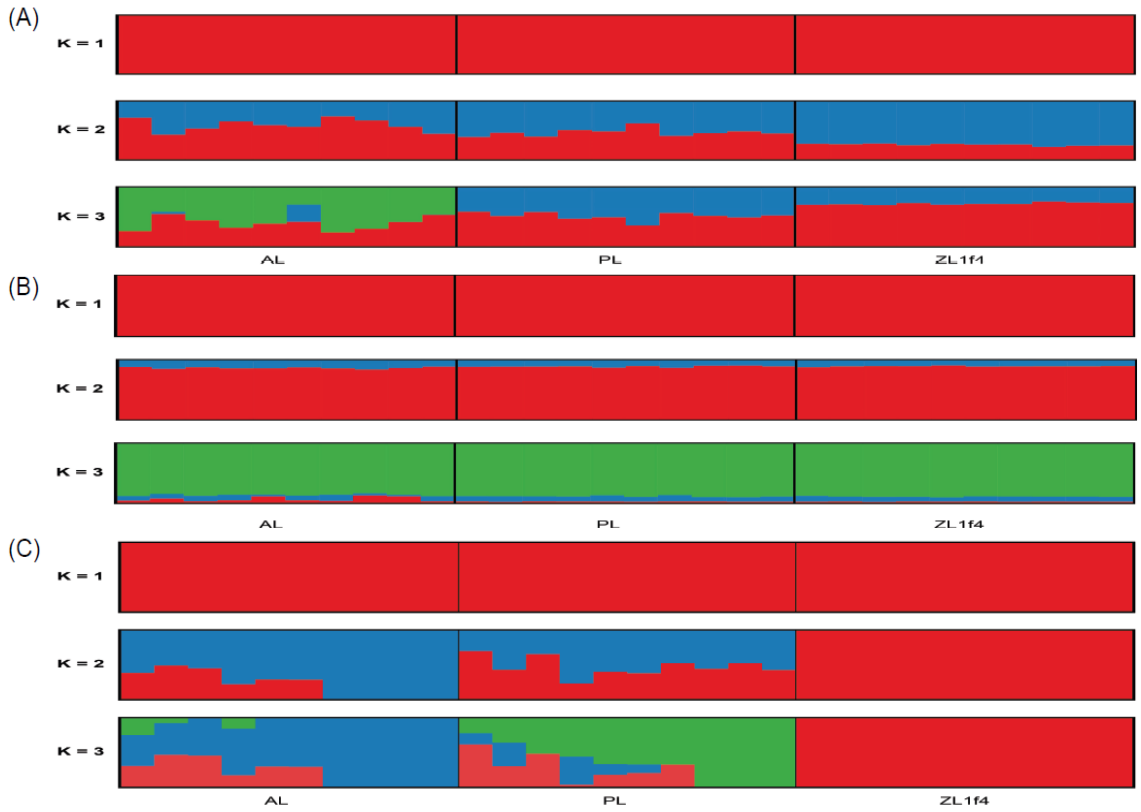

Figure 7. Genetic Structure of Largemouth Black Bass Populations, Assessed Using Different Marker Types for Inferred Genetic Clusters (K=1, K=2, and K=3). Plots for SSR markers (panels A and B) were analyzed by STRUCTURE, while SNP markers (panel C) were analyzed by ADMIXTURE. (A) Structure plots derived from LobSTR-genotyped SSR markers. (B) Structure plots derived from SSRgenotyper-genotyped SSR markers. (C) Structure plots derived from SNP markers. In all panels, each vertical bar represents an individual Largemouth Black Bass, and the colored segments within each bar indicate the proportion of ancestry derived from the inferred genetic clusters. Populations include Anhui (AL), Panzhihua (PL), and the selectively bred 'Zhejiang Black Bass No.1' F4 (ZL1f4).

comprehensive analysis highlights the differential sensitivity and specificity of the genotyping tools employed—LobSTR, SSRgenotyper, and SNP markers — in capturing various aspects of genetic diversity within populations. While LobSTR generally provided more precise estimates of genetic diversity compared to SSRgenotyper, a stronger correlation was observed between SSRgenotyper and SNP results, which is likely attributable to SSRgenotyper's direct identification and genotyping of SSR loci from resequencing data.

Genetic Diversity Trends and Aquaculture Implications

Our findings revealed significant disparities in genetic diversity metrics among the ZL1f4, AL, and PL populations. Genome-wide assessments consistently showed that genetic diversity indices such as *H*e and π followed a clear trend: AL>PL>ZL1f4. This pattern strongly suggests a reduction in genetic diversity within the ZL1f4 population, likely reflecting the impact of the selective breeding program. This observed reduction in genetic diversity is a critical finding with direct implications for aquaculture. A diminished genetic base can compromise a population's long-term adaptability to environmental changes, disease resistance, and overall productivity, potentially leading to inbreeding depression (Wilkins et al., 2014). To mitigate these risks and ensure the sustained improvement and resilience of Largemouth black bass in aquaculture, future breeding strategies should actively consider implementing practices such as diversifying broodstock sources, employing rotational breeding schemes, and carefully monitoring genetic load to maintain sufficient genetic variation. Our results underscore the vital role of genetic diversity in maintaining the adaptability and health of aquaculture populations, aligning with insights from previous research (Bai et al., 2008).

Elucidating Population Structure and Phylogenetic Relationships

Through PCoA and phylogenetic analysis, we further confirmed the genetic differences between the ZL1f4 and AL, PL populations. While the NJ tree, based on genetic distances, broadly distinguished between different populations, the support for these branches was often low, likely due to the minimal genetic distances within the species, consistent with previous findings (Bai et al., 2008). Our PCoA results, however, more effectively differentiated between populations, particularly when using SNPs, followed by LobSTR and SSRgenotyper. The superior resolution provided by SNP markers in PCoA is a key finding, demonstrating their efficacy in precisely delineating genetic relationships and the impact of artificial selection on population structure, which is crucial for effective breeding management. Phylogenetic networks provided complementary insights, confirming genetic relatedness while also revealing complex interconnections. The differences observed between SSRgenotyper and the other two marker types again underscore the importance of marker choice and genotyping methodology in shaping the perceived genetic landscape. The smaller confidence ellipse for the ZL1f4 population on the PCoA plot suggests a reduced variance in genetic distance within this population, likely attributable to artificial selection, aligning with similar trends reported by Sun et al. (2023).

Impact of Artificial Selection on Genetic Structure

Genetic structure analysis revealed that despite observed genetic differences, the ZL1f4, AL, and PL populations still largely share the same ancestral component. This suggests that artificial selection, while influencing genetic diversity, has not fundamentally altered the broader genetic structure of Largemouth black bass, which is a positive outcome for the conservation of genetic resources and future genetic improvement. Our findings indicate that artificial selection reduced genetic diversity in the ZL1f4 population and further confirmed genetic differences between the ZL1f4 and AL, PL populations through PCoA analysis. However, this did not result in significant changes in genetic structure. This may be because artificial selection primarily targets genes or genomic regions associated with specific traits rather than inducing genome-wide structural changes (Plassais et al., 2022; Wu et al., 2024). Our results also suggest that precise selection and management can optimize traits without disrupting the fundamental genetic structure. It is crucial to highlight that our findings on the genetic structure consistency among the ZL1f4, AL, and PL populations diverge from some previous studies (Su et al., 2020; Sun et al., 2023; Zhang et al., 2022), which did not consistently evaluate whether K=1 was the optimal number of genetic clusters (K value). Our analysis using LobSTR and SSRgenotyper suggests that K=2 is often the optimal value, and Figure 7 illustrates the striking similarity in genetic structures across the ZL1f4, AL, and PL populations. We posit that the abbreviated selection duration for the ZL1f4 population has not yet markedly altered its genetic structure, a situation potentially tied to specific breeding methods and management practices. This aligns with observations in other species where intensive artificial selection can modify genetic architecture without completely erasing ancestral patterns (Naval-Sanchez et al., 2020; Seo et al., 2022).

Comparative Efficacy and Limitations of Whole-genome SSR and SNP Markers

Historically, the labor-intensive, costly, and time-consuming nature of traditional SSR genotyping methods has limited the use of extensive SSR markers in genetic diversity studies. However, advancements in sequencing technologies have made whole-genome SSR genotyping a more attractive and feasible alternative.

Our preliminary studies suggest that whole-genome SSR genotyping can identify a significantly larger number of polymorphic SSR markers (Liu, Xie, 2022; Liu et al., 2023c), offering fresh insights into previous research (Liu et al., 2023a). Our comparisons between LobSTR, SSRgenotyper, and SNP analyses consistently revealed inherent differences in the results obtained by the different marker types and genotyping tools. This discrepancy highlights the unique characteristics of whole-genome SSR markers relative to whole-genome SNP markers. For instance, Han et al. (2022) demonstrated that genetic analyses based on whole-genome SSR markers yielded results distinct from those obtained using whole-genome SNP markers, indicating the divergent insights these technologies can offer into genetic diversity. This finding is consistent with other research (Chen et al., 2022; Liu et al., 2023a), emphasizing that the choice of molecular markers can significantly impact results across different research contexts. While SNPs possess inherent strengths due to their abundance and uniform genomic distribution, SSRs offer unique advantages, such as higher mutation rates and multi-allelic nature, that are not easily replaceable (Hauser et al., 2021). This underscores the concept of SSR markers serving as a valuable complement to SNPs in the toolkit for comprehensive genetic analysis. However, it is crucial to acknowledge the limitations of each marker type and genotyping approach. For instance, while SSRs offer high polymorphism, their accurate genotyping from whole-genome resequencing data can be challenging, as evidenced by the lower accuracy and higher rate of missing data associated with tools like SSRgenotyper. Conversely, while SNPs provide extensive genomic coverage, their genotyping can be susceptible to biases introduced by the choice of reference genome or specific filtering methodologies. To ensure the accuracy of SSR analysis outcomes, including a sufficient number of high-quality SSRs in genetic diversity studies is essential (Liu et al., 2023a).

For SNP analysis, loci devoid of strong linkage disequilibrium were selected via Plink v1.9, applying the settings --indep-pairwise 100 10 0.5 prior to genetic structure analysis (Admixture). This step ensures that the SNPs used for population structure inference are largely independent, preventing overestimation of genetic differentiation due to linked markers. In contrast, LD pruning was not performed for SSR markers for genetic variation and diversity analysis. This decision was based on the inherent characteristics of SSRs, which are typically highly polymorphic and multi-allelic, and are often used to capture more recent evolutionary events or fine-scale population structure where strong LD across large genomic regions is less of a primary concern compared to biallelic SNPs (Chen et al., 2024; Han et al., 2022). Furthermore, standard LD pruning methodologies developed for biallelic SNPs are not directly transferable or commonly applied to multi-allelic SSR data in the same manner. Therefore, all identified polymorphic SSR loci (28,250 for LobSTR and 2,142 for SSRgenotyper) were utilized in the subsequent genetic diversity and structure analyses for their respective datasets.

Influence of Genotyping Tools and Reference Genomes on Diversity Estimates

Our study revealed that the genetic diversity indices derived from SSRgenotyper analysis were consistently lower than those reported in prior research (Bai et al., 2008; Fu et al., 2023; Liang et al., 2008; Su et al., 2020; Sun et al., 2019; Wang et al., 2019; Wang et al., 2020; Zhang et al., 2022). In contrast, indices from LobSTR aligned closely with earlier findings, such as the average *Ho* for the PL population (0.509 in this study vs. 0.508 reported by Bai et al. (2008) for wild populations). This highlights the significant influence of genotyping tools on the interpretation of genetic diversity and the assessment of genetic distances within populations. The discrepancies noted with SSRgenotyper likely stem from its specific method of directly identifying and genotyping SSR loci from resequencing data, which might explain its higher correlation of genetic distances with SNPs despite its lower diversity estimates. Whole-genome resequencing remains a complex challenge for accurate SSR genotyping, with LobSTR being one of the pioneering successful tools. While other tools like HipSTR (Willems et al., 2017) have gained popularity, our choice of LobSTR was based on its favorable error rate (slightly under 1%) and faster genotyping speed (Halman, Oshlack, 2020). Although SSRgenotyper exhibited significantly lower estimated genetic diversity indices, its ability to genotype without a reference genome makes it suitable for cross-species SSR genotyping, particularly for species lacking reference genomes (Liu et al., 2023a). However, its limitations in accuracy and higher rate of missing data underscore the trade-offs inherent in different genotyping principles.

Furthermore, our analysis of whole-genome SNP markers indicated that π between the selected breeding and introduced cultured populations followed the trend AL>PL>ZL1f4, with an average of 1.17×10^{-3}. While higher than results from some similar Largemouth black bass studies (Du et al., 2022; Sun et al., 2023), this value was surprisingly lower than some critically endangered or vulnerable fish species listed by the IUCN (Galland et al., 2021; Pujolar et al., 2013). This discrepancy serves as a critical reminder that the reduction in genetic diversity in Largemouth black bass is a significant issue and a concern for germplasm improvement, emphasizing the need for ongoing monitoring and strategies to enhance genetic diversity. The observed discrepancies with prior studies, such as Sun et al. (2023), are likely multifaceted. While sampling errors and variations in SNP marker screening methodologies could contribute, we attribute the primary cause to the use of different reference genomes. Our study utilizes the chromosome-level, more contiguous reference genome reported by He et al. (2022), whereas Sun et al. (2023) employed an earlier

reference genome (Sun et al., 2021). The selection of the reference genome critically influences the identification of whole-genome SNP markers, and the use of disparate genomes can introduce significant biases (Liu et al., 2023b), a phenomenon extensively documented in the literature (Brandt et al., 2015; Gunther, Nettelblad, 2019; Stevenson et al., 2013). These differences in genomic resources and bioinformatic pipelines can lead to variations in identified polymorphic sites and subsequent diversity estimates, highlighting the importance of standardized approaches for robust comparisons.

Conclusion

In summary, our study elucidates the differences between whole-genome SSR and SNP markers in genetic analysis, demonstrating their complementary roles in providing comprehensive insights into genetic diversity and population structure. LobSTR generally delivered more precise genotyping than SSRgenotyper, yet the strong correlation between SSRgenotyper and SNP results underscores the necessity of using multiple markers for comprehensive evaluations, while also acknowledging the specific limitations of each genotyping approach. The study also offers valuable insights into Largemouth black bass's genetic enhancement and population management. It reveals significant differences in genetic diversity between the selectively bred population (ZL1f4) and the introduced cultured populations (AL, PL) while showing a fundamental consistency in genetic structure. This reflects the inherent genetic characteristics of the northern subspecies of Largemouth black bass. Our findings deepen the understanding of artificial selection's impact on genetic diversity and structure, guiding genetic improvement and biodiversity conservation in aquaculture. Additionally, our research introduces new perspectives for genetic studies on other domesticated species, contributing to the sustainable development of aquaculture. However, it is important to reiterate that limitations such as a relatively small sample size and the exclusion of environmental factors in the current analysis may affect the accuracy of genetic diversity estimates. Future research should aim to increase sample sizes and incorporate environmental influences, and potentially explore genomic regions under selection to provide more accurate and nuanced assessments of genetic diversity and adaptation

Ethical Statement

All experimental procedures involving fish were conducted in accordance with the guidelines established by the Institute of Fishery Science at the Hangzhou Academy of Agricultural Sciences and the Zhejiang Fisheries Technical Extension Center. The study protocol was approved by the administrative committee of both institutions. The fish species utilized in this research are not classified as rare or endangered, nor are they listed under China's first- or second-class state protection levels. All procedures were performed in strict adherence to the regulations stipulated by China's Wildlife Protection Law and Fishery Law.

Funding Information

This work was supported by the Key Scientific and Technological Grant of Zhejiang for Breeding New Agricultural (Aquaculture) Varieties [Grant numbers 2021C02069-2].

Author Contribution

Kai Liu: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization and Writing - original draft; Qinghui Meng: Investigation, Data Curation, Methodology, Writing -review and editing; Wenrui Sh: Data Curation, Resources; and Nan Xie: Data Curation, Resources.

Conflict of Interest

The author(s) declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

References

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 12, 246. https://doi.org/10.1186/1471-2105-12-246

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655-1664. https://doi.org/10.1101/gr.094052.109

Archer, F.I., Adams, P.E., Schneiders, B.B., 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources. 17, 5-11. https://doi.org/10.1111/1755-0998.12559

Bai, J., Lutz-Carrillo, D.J., Quan, Y., Liang, S., 2008. Taxonomic status and genetic diversity of cultured largemouth bass micropterus salmoides in China. Aquaculture. 278, 27-30. https://doi.org/10.1016/j.aquaculture.2008.03.016

Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., Ramachandran, S., 2016. Pong: Fast analysis and visualization of latent clusters in population genetic data. Bioinformatics. 32, 2817-2823. https://doi.org/10.1093/bioinformatics/btw327

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573-580. https://doi.org/10.1093/nar/27.2.573

Brandt, D.Y., Aguiar, V.R., Bitarello, B.D., Nunes, K., Goudet, J., Meyer, D., 2015. Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. G3 (Bethesda). 5, 931-941.https://doi.org/10.1534/g3.114.015784.

Broad, 2019. Picard toolkit. https://broadinstitute.github-.io/picard/. Accessed December 10th, 2021.

Chen, J.-A., Yu, P.-J., Jheng, S.-W., Lin, Y.-Z., Sun, P.-W., Ko, W.-Y., Lin, C.-F., Ju, Y.-T., 2024. Mining expressed sequence tag (EST) microsatellite markers to assess the genetic differentiation of five hynobius species endemic to taiwan. Scientific Reports. 14, 20898. https://doi.org/10.1038/s41598-024-71887-1

Chen, J., Liu, H., Gooneratne, R., Wang, Y., Wang, W., 2022. Population Genomics of Megalobrama Provides Insights into Evolutionary History and Dietary Adaptation. Biology (Basel). 11, 186. https://doi.org/10.3390/biology11020186

Chen, S., 2023. Ultrafast one‐pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2. https://doi.org/10.1002/imt2.107

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10, 1-4. https://doi.org/10.1093/gigascience/giab008

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., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. Bioinformatics. 27, 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43, 491-498. https://doi.org/10.1038/ng.806

Du, J., Li, S., Shao, J., Song, H., Jiang, P., Lei, C., Bai, J., Han, L., 2022. Genetic diversity analysis and development of molecular markers for the identification of largemouth bass (Micropterus salmoides L.) based on whole-genome re-sequencing. Front Genet. 13, 936610. https://doi.org/10.3389/fgene.2022.936610

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

Fu, J.-j., Gong, Y.-t., Zhu, W.-b., Qiang, J., Zhang, L.-b., Wang, L.-m., Luo, M.-k., Dong, Z.-j., 2023. Multiplex pcr sets of microsatellites and their application in genetic analyses for three populations of largemouth bass (micropterus salmoides). Acta Hydrobiologica Sinica. 47, 1514-1522.

Galland, L.M., Simmons, J.B., Jahner, J.P., Luzuriaga-Neira, A.R., Sloat, M.R., Chandra, S., Hogan, Z., Jensen, O.P., Parchman, T.L., 2021. Hierarchical genetic structure and implications for conservation of the world's largest salmonid, hucho taimen. Scientific Reports. 11, 20508. https://doi.org/10.1038/s41598-021-99530-3

Green, M.R., Sambrook, J., 2018. Isolation of High-Molecular-Weight DNA from Suspension Cultures of Mammalian Cells Using Proteinase K and Phenol. Cold Spring Harb Protoc. 2018, 317-321. https://doi.org/10.1101/pdb.prot093476

Gunther, T., Nettelblad, C., 2019. The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. 15, e1008302. https://doi.org/10.1371/journal.pgen.1008302

Guo, J., Zhang, M., Wang, S., Xu, X., Shen, Y., Li, J., 2022. A high-density genetic linkage map and QTL mapping for growth related traits in grass carp (ctenopharyngodon idella). Aquaculture. 552, 738041. https://doi.org/10.1016/j.aquaculture.2022.738041

Gymrek, M., Golan, D., Rosset, S., Erlich, Y., 2012. lobSTR: A short tandem repeat profiler for personal genomes. Genome Res. 22, 1154-1162. https://doi.org/10.1101/gr.135780.111

Halman, A., Oshlack, A., 2020. Accuracy of short tandem repeats genotyping tools in whole exome sequencing data. F1000Research. 9. https://doi.org/10.12688/f1000research.22639.1.

Han, J., Munro, J.E., Kocoski, A., Barry, A.E., Bahlo, M., 2022. Population-level genome-wide STR discovery and validation for population structure and genetic diversity assessment of Plasmodium species. PLoS Genet. 18, e1009604. https://doi.org/10.1371/journal.pgen.1009604

Hauser, S.S., Athrey, G., Leberg, P.L., 2021. Waste not, want not: Microsatellites remain an economical and informative technology for conservation genetics. Ecol Evol. 11, 15800-15814. https://doi.org/10.1002/ece3.8250

He, K., Zhao, L., Yuan, Z., Canario, A., Liu, Q., Chen, S., Guo, J., Luo, W., Yan, H., Zhang, D., Li, L., Yang, S., 2022. Chromosome-level genome assembly of largemouth bass (Micropterus salmoides) using PacBio and Hi-C technologies. Sci Data. 9, 482. https://doi.org/10.1038/s41597-022-01601-1

Hodel, R.G., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A., Douglas, N.A., Germain-Aubrey, C.C., Chen, S., Soltis, D.E., Soltis, P.S., 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. Appl Plant Sci. 4. https://doi.org/10.3732/apps.1600025

Huang, X., Jiang, Y., Zhang, W., Cheng, Y., Wang, Y., Ma, X., Duan, Y., Xia, L., Chen, Y., Wu, N., Shi, M., Xia, X.-Q., 2020. Construction of a high-density genetic map and mapping of growth related QTLs in the grass carp (ctenopharyngodon idellus). BMC Genomics. 21, 313. https://doi.org/10.1186/s12864-020-6730-x.

Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour. 9, 1322-1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x.

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23, 1801-1806. https://doi.org/10.1093/bioinformatics/btm233.

Jombart, T., Ahmed, I., 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics. 27, 3070-3071. https://doi.org/10.1093/bioinformatics/btr521.

Jung, Y., Han, D., 2022. BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics. 38, 2404-2413. https://doi.org/10.1093/bioinformatics/btac137.

Kamvar, Z.N., Brooks, J.C., GrÃ¼nwald, N.J., 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics. 6. https://doi.org/10.3389/fgene.2015.00208.

Knaus, B.J., Grünwald, N.J., 2017. Vcfr: A package to manipulate and visualize variant call format data in R. Molecular Ecology Resources. 17, 44-53. https://doi.org/10.1111/1755-0998.12549.

Lewis, D.H., Jarvis, D.E., Maughan, P.J., 2020. SSRgenotyper: A simple sequence repeat genotyping application for whole-genome resequencing and reduced representational sequencing projects. Appl Plant Sci. 8, e11402. https://doi.org/10.1002/aps3.11402.

Liang, S.X., Sun, X.W., Bai, J.J., Gao, J.S., 2008. Genetic analysis for cultured largemouth bass (Micropterus salmoides) in China with microsatellites. Acta Hydrobiologica Sinica, 694-700.

Liu, K., Xie, N., 2022. Pipeline for developing polymorphic microsatellites in species without reference genomes. 3 Biotech. 12, 248. https://doi.org/10.1007/s13205-022-03313-0

Liu, K., Xie, N., Wang, Y., 2023a. Exploring cross-species genetic diversity: unveiling new insights in Megalobrama through whole genome-wide simple sequence repeats. Conservation Genetics. 25, 393-407. https://doi.org/10.1007/s10592-023-01575-6

Liu, K., Xie, N., Wang, Y., Liu, X., 2023b. Contribution bias of parental genomes to the hybrid lineages of black Amur bream and topmouth culter revealed by low-coverage whole-genome sequencing. Gene. 852, 147058. https://doi.org/10.1016/j.gene.2022.147058

Liu, K., Xie, N., Wang, Y., Liu, X., 2023c. Development of microsatellite markers in Macrobrachium nipponensebased on sequencing genotyping. Journal of Fisheries Research. 45, 537-543. https://doi.org/10.14012/j.cnki.fjsc.2023.06.003

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research. 27, 209-220.

Mijangos, J.L., Gruber, B., Berry, O., Pacioni, C., Georges, A., 2022. dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210x.13918

Naval-Sanchez, M., McWilliam, S., Evans, B., Yáñez, J.M., Houston, R.D., Kijas, J.W., 2020. Changed Patterns of Genomic Variation Following Recent Domestication: Selection Sweeps in Farmed Atlantic Salmon. Frontiers in Genetics. 11. https://doi.org/10.3389/fgene.2020.00264

Nei, M., 1972. Genetic Distance between Populations. The American Naturalist. 106, 283-292. https://doi.org/10.1086/282771

Paradis, E., 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics. 26, 419-420. https://doi.org/10.1093/bioinformatics/btp696

Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources. 13, 946-952. https://doi.org/10.1111/1755-0998.12129

Plassais, J., vonHoldt, B.M., Parker, H.G., Carmagnini, A., Dubos, N., Papa, I., Bevant, K., Derrien, T., Hennelly, L.M., Whitaker, D.T., Harris, A.C., Hogan, A.N., Huson, H.J., Zaibert, V.F., Linderholm, A., Haile, J., Fest, T., Habib, B., Sacks, B.N., Benecke, N., Outram, A.K., Sablin, M.V., Germonpré, M., Larson, G., Frantz, L., Ostrander, E.A., 2022. Natural and human-driven selection of a single non-coding body size variant in ancient and modern canids. Current Biology. 32, 889-897.e889. https://doi.org/10.1016/j.cub.2021.12.036

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics. 155, 945-959. https://doi.org/10.1093/genetics/155.2.945

Pujolar, J.M., Jacobsen, M.W., Frydenberg, J., Als, T.D., Larsen, P.F., Maes, G.E., Zane, L., Jian, J.B., Cheng, L., Hansen, M.M., 2013. A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. Mol Ecol Resour. 13, 706-714. https://doi.org/10.1111/1755-0998.12117

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W.d., Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 81, 559-575. https://doi.org/10.1086/519795

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J., 2008. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 5, 1005-1010. https://doi.org/10.1038/nmeth.1270

Rohlf, F.J., Sokal, R.R., 1995. Biometry: The principles and practice of statistics in biological research, 3rd Editon ed. W. H. Freeman and Company, New York.

Rosenberg, N.A., 2003. distruct: a program for the graphical display of population structure. Mol Ecol Notes. 4, 137-138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Seo, D., Lee, D.H., Jin, S., Won, J.I., Lim, D., Park, M., Kim, T.H., Lee, H.K., Kim, S., Choi, I., Lee, J.H., Gondro, C., Lee, S.H., 2022. Long-term artificial selection of hanwoo (korean) cattle left genetic signatures for the breeding traits and has altered the genomic structure. Scientific Reports. 12, 6438. https://doi.org/10.1038/s41598-022-09425-0

Stevenson, K.R., Coolon, J.D., Wittkopp, P.J., 2013. Sources of bias in measures of allele-specific expression derived from RNA-sequence data aligned to a single reference genome. BMC Genomics. 14, 536. https://doi.org/10.1186/1471-2164-14-536.

Su, S., Zhang, L., Li, H., Gao, C., He, X., Tian, C., Li, J., Wang, M., Tang, Y., 2020. Genetic diversity and structure analyses of largemouth bass (Micropterus salmoides) original and cultured populations based on microsatellite markers. Journal of Zhejiang University (Agriculture and Life Sciences). 46, 687-698.

Sun, C., Xie, W., Hu, J., Dong, J., Tian, Y., Wu, Z., Ye, X., 2019. Genetic diversity analysis of three cultured populations of Micropterussalmoides. South China Fisheries Science. 15, 64-71.

Sun, C., Li, J., Dong, J., Niu, Y., Hu, J., Lian, J., Li, W., Li, J., Tian, Y., Shi, Q., Ye, X., 2021. Chromosome-level genome assembly for the largemouth bass Micropterus salmoides provides insights into adaptation to fresh and brackish water. Mol Ecol Resour. 21, 301-315. https://doi.org/10.1111/1755-0998.13256

Sun, C.F., Zhang, X.H., Dong, J.J., You, X.X., Tian, Y.Y., Gao, F.Y., Zhang, H.T., Shi, Q., Ye, X., 2023. Whole-genome resequencing reveals recent signatures of selection in five populations of largemouth bass (Micropterus salmoides). Zool Res. 44, 78-89. https://doi.org/10.24272/j.issn.2095-8137.2022.274

Szatmari, L., Cserkesz, T., Laczko, L., Lanszki, J., Pertoldi, C., Abramov, A.V., Elmeros, M., Ottlecz, B., Hegyeli, Z., Sramko, G., 2021. A comparison of microsatellites and genome-wide SNPs for the detection of admixture brings the first molecular evidence for hybridization between Mustela eversmanii and M. putorius (Mustelidae, Carnivora). Evol Appl. 14, 2286-2304. https://doi.org/10.1111/eva.13291

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2015. Sambamba: Fast processing of NGS alignment formats. Bioinformatics. 31, 2032-2034. https://doi.org/10.1093/bioinformatics/btv098

Thiel, T., Michalek, W., Varshney, R.K., Graner, A., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 106, 411-422. https://doi.org/10.1007/s00122-002-1031-0

Tsykun, T., Rellstab, C., Dutech, C., Sipos, G., Prospero, S., 2017. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus Armillaria cepistipes. Heredity (Edinb). 119, 371-380. https://doi.org/10.1038/hdy.2017.48

Wang, D., Yao, H., Li, Y.H., Xu, Y.J., Ma, X.F., Wang, H.P., 2019. Global diversity and genetic landscape of natural populations and hatchery stocks of largemouth bass micropterus salmoides across American and Asian regions. Sci Rep. 9, 16697. https://doi.org/10.1038/s41598-019-53026-3.

Wang, J., 2019. A parsimony estimator of the number of populations from a STRUCTURE-like analysis. Mol Ecol Resour. 19, 970-981. https://doi.org/10.1111/1755-0998.13000

Wang, P., Zhou, G., Chen, S., Lu, J., 2020. Analysis of growth trait comparison and genetic diversity of F1 progeny on cross species of southern largemouth bass, northern largemouth bass and "Youlu No. 3". Marine Fisheries. 42, 403-409. https://doi.org/10.13233/j.cnki.mar.fish.2020.04.003.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 38. https://doi.org/10.2307/2408641

Wilkins, A.S., Wrangham, R.W., Fitch, W.T., 2014. The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. Genetics. 197, 795-808. https://doi.org/10.1534/genetics.114.165423

Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., Erlich, Y., 2017. Genome-wide profiling of heritable and de novo STR variations. Nat Methods. 14, 590-592. https://doi.org/10.1038/nmeth.4267

Wright, S., 1978. Genetic variability in natural populations: methods, Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. University of Chicagio Press, Chicagio.

Wu, M., Wang, D., Li, M.-H., Lv, F., 2024. Artificial selection shapes the lower genomic diversity and higher selective pressures on the sex chromosomes of domestic animals. Sci. China Life Sci. https://doi.org/10.1007/s11427-023-2478-5

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.Y., McInerny, G., 2016. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 8, 28-36. https://doi.org/10.1111/2041-210x.12628.

Zhang, D., Qiang, J., Fu, J., Xu, G., Zhou, G., Dong, Z., Tao, Y., Lu, S., Lu, J., Xu, P., 2022. Genetic analysis of five stocks of largemouth bass (Micropterus salmoides) with microsatellite and mitochondrial D-loop sequences. Journal of Fishery Sciences of China. 29, 1277-1289.

Zhang, Q., Chen, J., Jiang, X., Zou, S., 2014. Establishment of DNA fingerprinting and analysis on genetic structure of different Parabramis and Megalobrama populations with microsatellite. Journal of Fisheries of China. 38, 15-22.

Zimmerman, S.J., Aldridge, C.L., Oyler-McCance, S.J., 2020. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. BMC Genomics. 21, 382. https://doi.org/10.1186/s12864-020-06783-9