

## Original Research Articles

# A parentage identification system and genetic analysis of growth traits in *Percocypris pingi* (Tchang, 1930) for breeding programs

Siyi Zhu<sup>1</sup>, Yun Dai<sup>1</sup>, Longjun Deng<sup>2</sup>, Tiancai Li<sup>2</sup>, Tong Li<sup>1</sup>, Qingjie Wang<sup>1</sup>, Shun Zhong, Xing Ying<sup>1</sup>, Fulin Dai<sup>1</sup>, Dongjie Wang<sup>1</sup>, Wei Luo<sup>1a</sup>, Zongjun Du<sup>1b</sup>

<sup>1</sup> College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China, <sup>2</sup> Yalong River Hydropower Development Co., Ltd., Chengdu 610051, China

Keywords: Stock enhancement, parentage identification, genetic diversity, heritability, growth traits, *Percocypris pingi*

<https://doi.org/10.46989/001c.154808>

## Israeli Journal of Aquaculture - Bamidgheh

Vol. 77, Issue 4, 2025

To effectively evaluate genetic diversity and estimate trait heritability, the establishment of a robust parentage identification system is indispensable. For tetraploid species, the success of such a system critically depends on the development of highly polymorphic loci. In this study, we screened simple sequence repeats (SSRs) for *Percocypris pingi*, a tetraploid species, by leveraging full-length data. Using SSREnricher software, we further compared these sequences with muscle transcriptome data obtained from four geographically distinct populations to identify highly polymorphic SSR loci. We identified nine highly polymorphic SSR markers with a mean polymorphism information content (PIC) of 0.647 and constructed a high-resolution parentage assignment system with a cumulative exclusion probability of 0.999984. The system was validated using 300 offspring from single-pair families, achieving an assignment accuracy of 96%. Growth traits, including total length (TL), body length (BL), and body weight (BW), were measured in 270 individuals from mixed-rearing groups. We investigated the relationship between parental genetic distance and offspring growth. The average Bruvo distance between parents was relatively low (0.37). A weak positive trend was observed between parental genetic distance and offspring TL ( $P > 0.05$ ). Given that all pairwise trait correlations exceeded 0.9 ( $P < 0.05$ ), the multivariate model provided more reliable estimates, and BW (heritability = 0.307) emerged as a priority trait for future breeding programs. Overall, our findings highlight the utility of transcriptome-derived SSR markers for accurate parentage analysis and emphasize the importance of monitoring genetic diversity and estimating trait heritability in stock enhancement efforts.

## INTRODUCTION

Stock enhancement, a pivotal strategy for fisheries restoration, has long been a subject of extensive debate. Extensive research has demonstrated that the widespread release of hatchery-reared fish into the natural environments can introduce significant genetic and ecological risks.<sup>1,2</sup> These risks are primarily linked to declines in genetic diversity within hatchery populations.<sup>3</sup> A marked reduction in genetic diversity can lead to severe consequences, such as diminished adaptive potential, increased homozygosity, altered phenotypes, and reduced overall fitness.<sup>4</sup> To address these risks, hatcheries commonly source broodstock directly from wild populations rather than relying on successive generations of hatchery-born individuals. Nevertheless, ensuring high levels of genetic diversity among

offspring remains an ongoing challenge. Two primary factors contribute to this issue: the limited availability of wild broodstock and a lower-than-expected number of effective breeders, both of which result in reduced genetic variation in offspring populations.

The selection of unrelated individuals for broodstock formation is widely recognized as a robust strategy for maintaining genetic diversity.<sup>5</sup> Precise kinship estimation plays a critical role in facilitating informed decision-making during broodstock selection. Over the past few years, microsatellite markers have demonstrated remarkable efficacy in kinship analysis, providing high accuracy even with a limited number of loci. The advancement of multiplex PCR techniques, which enable the simultaneous amplification of multiple microsatellite loci, has significantly enhanced both the efficiency and cost-effectiveness of kinship

<sup>a</sup> Corresponding author, Wei Luo, e-mail: [14364@sicau.edu.cn](mailto:14364@sicau.edu.cn)

<sup>b</sup> Corresponding author, Zongjun Du, e-mail: [10988@sicau.edu.cn](mailto:10988@sicau.edu.cn)

analysis., These methodologies have been successfully implemented across variety of aquaculture species.<sup>6</sup> Furthermore, empirical investigations indicate that minimum kinship mating schemes are more effective at preserving allele diversity and heterozygosity compared to random mating, thereby underscoring the importance of kinship estimation as a key tool in genetic conservation initiatives.

Understanding the genetic regulation of productive traits in fish is critically important, as these traits frequently exhibit substantial genetic variation and are directly associated with economic value. Through the application of molecular parentage analysis, more precise estimations of genetic parameters can be attained, even within mixed-rearing systems where pedigree tracking poses significant challenges.<sup>7</sup> Consequently, enhancing fish production efficiency via selective breeding and genetic improvement of broodstock populations represents a rational and viable approach, which has been successfully applied across various aquaculture species.<sup>8</sup>

The population of *Percocypris pingi* (Tchang, 1930) has undergone a significant decline as a result of dam construction and overfishing, prompting its inclusion on both the IUCN Red List of Threatened Species and China's List of National Key Protected Wild Animals. Studies on the genetic diversity of cultured and wild *P. pingi* populations in the Yalong River have revealed moderate levels of observed heterozygosity (0.657–0.770).<sup>9</sup> While some SSR markers have been developed for this species—such as 12 tetranucleotide markers by Deng et al.<sup>10</sup> and 20 markers identified through transcriptome sequencing scanning by Wu et al.<sup>11</sup> their utility in parentage analysis remains restricted. Additionally, as *P. pingi* is a tetraploid species, the development of highly polymorphic SSR loci is critical for the success of genetic analyses. To support conservation and stock enhancement initiatives, this study utilized the full-length transcriptome of *P. pingi* (CRA016173), complemented by additional transcriptome data (CRA024585), to identify polymorphic SSR loci and establish a robust parentage identification system. The accuracy of this system was validated using single-pair families, while the relationship between genetic distance and growth performance, as well as the estimation of genetic parameters, was investigated using mixed-rearing groups. The correlation between parental genetic distance (measured by Bruvo's distance) and offspring growth performance was assessed, and genetic parameter estimation was conducted using both univariate and multivariate animal models (Fig. 1). This research aims to provide a theoretical basis and technical framework for future breeding and conservation programs targeting *P. pingi*.

## MATERIALS AND METHODS

### 2.1. EXPERIMENTAL FISH AND EXPERIMENTAL DESIGN

This study involved three groups of *P. pingi*: (1) broodstock utilized for breeding purposes, (2) individuals sampled from diverse geographic locations for transcriptome sequencing

and development of polymorphic SSR markers, and (3) offspring bred to validate the breeding strategy.

#### 2.1.1. BROODSTOCK REARING AND MANAGEMENT

Healthy, disease-free, and sexually mature *P. pingi* broodstock were carefully selected from the Yalong River Jinping–Guandi Fish Stock Enhancement and Release Station and subsequently transferred to a broodstock conditioning tank for temporary rearing. A recirculating aquaculture system (RAS) was employed to ensure stable and optimal water quality conditions, including temperature (12–22°C), dissolved oxygen (DO > 8 mg/L), ammonia (< 0.1 mg/L), nitrite (< 0.01 mg/L). Following a seven-day acclimation period, six females (mean total length = 59.46 ± 6.92 cm; mean body weight = 2722.93 ± 533.59 g) and six males with well-developed gonads (mean total length = 46.19 ± 4.15 cm; mean body weight = 1092.35 ± 282.53 g) were chosen as broodstock. During artificial fertilization, the total length (TL), standard length (SL), and body weight (BW) of all broodstock individuals were meticulously recorded (Table S1). Caudal fin samples were collected from each individual, preserved in ethanol, and stored at –20°C for subsequent DNA extraction.

#### 2.1.2. FRY HATCHING AND JUVENILE MANAGEMENT

Based on a balanced nested design, 36 full-sib families were initially planned for establishment (Table S2). Fertilized eggs were transported to the Aquaculture Laboratory of Sichuan Agricultural University for hatching and rearing. Post-hatching, larvae initiated free-swimming within 5–7 days. Once all larvae exhibited free-swimming capability, 100 larvae from each *P. pingi* family were randomly selected and placed into separate rearing frames within the same nursery tank, forming the single-pair families. Concurrently, another 100 larvae from each family were randomly selected and communally reared in the same nursery tank, forming the mixed-rearing group.

The single-pair families were utilized for the development of a parentage identification technique, while the communal-rearing group served to validate this method and evaluate the genetic parameters of growth traits. During the rearing period, water temperature followed natural fluctuations, pH was maintained between 8.0 and 8.3, DO levels exceeded 8 mg/L, and both ammonia nitrogen and nitrite nitrogen concentrations remained below 0.01 mg/L. Fish were fed commercial feed twice daily, with uneaten feed and feces removed by siphoning one hour after each feeding.

After 30 days of rearing, 14 full-sib families of *P. pingi* with normal growth were successfully established in the single-pair families (Table S3). After 12 months of rearing, 300 juvenile fish were randomly selected from the 14 single-pair families for parentage identification accuracy verification. Additionally, 270 juvenile fish were randomly selected from the mixed-rearing group as experimental subjects. The TL (measured to the nearest 0.1 cm), SL (measured to the nearest 0.1 cm), and BW (measured to the nearest 0.1 g) of the offspring were recorded, and caudal fin

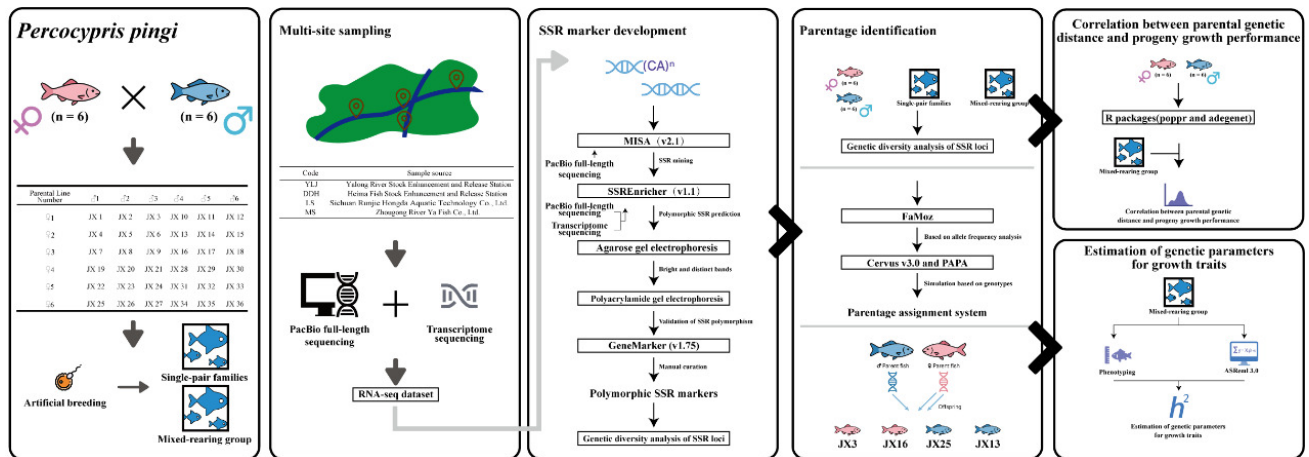


Fig. 1. Workflow diagram illustrating the experimental design and analysis pipeline.

samples were collected for parentage identification analysis.

## 2.2. FULL-LENGTH TRANSCRIPTOME SEQUENCING AND TRANSCRIPTOME SEQUENCING

### 2.2.1. COLLECTION OF FISH FOR TRANSCRIPTOME SEQUENCING

Individuals utilized for transcriptome sequencing were sourced from four locations: the Yalong River Jinping–Guandi Fish Stock Enhancement and Release Station, the Heima Fish Stock Enhancement and Release Station, Sichuan Runjie Hongda Aquatic Technology Co., Ltd., and Ya'an Zhougong River Yafish Co., Ltd. Full-length transcriptome sequencing was carried out on samples obtained from the Yalong River Stock Enhancement and Release Station, whereas throughput transcriptome sequencing was performed using individuals collected from all four locations (Table S4).

### 2.2.2. FULL-LENGTH TRANSCRIPTOME SEQUENCING

Full-length transcriptome sequencing was conducted on *P. pingi* specimens sourced from the Yalong River Jinping–Guandi Fish Stock Enhancement and Release Station. Juvenile fish, with an average weight of 9.43 g, were initially anesthetized using MS-222 at a concentration of 100 mg/L and subsequently dissected under sterile conditions. Tissue samples were harvested from multiple organs, including scales, skin, dorsal muscle, brain, gills, liver, kidney, spleen, and intestine (Table S5). All samples were transferred to pre-labeled cryovials, immediately flash-frozen in liquid nitrogen, and stored on dry ice prior to shipment to Beijing Biomarker Technologies Co., Ltd. (<http://www.biomarker.com.cn/>) for sequencing.

Total RNA extraction from each tissue was performed using the TRIzol reagent method as per the manufacturer's instructions. The RNA concentration and purity were quantified using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA) by assessing the A260/A280 ratio, while RNA integrity was verified using an Agilent 2100 Bio-

analyzer (Agilent Technologies, USA). For full-length transcriptome sequencing, cDNA libraries were constructed as follows: messenger RNA was reverse-transcribed into cDNA using the Clontech SMARTer PCR cDNA Synthesis Kit (Takara Bio, Japan), followed by PCR amplification to generate double-stranded cDNA. The cDNA underwent damage repair, end repair, and adapter ligation, with SMRT dumbbell adapters (Pacific Biosciences, Menlo Park, CA, USA) being ligated to produce SMRTbell libraries compatible with the PacBio Sequel II platform. Library quality was evaluated using both the Qubit 2.0 fluorometer (Invitrogen, USA) with the dsDNA HS Assay Kit and the Agilent 2100 Bioanalyzer.

High-fidelity circular consensus sequences (CCSs) were generated from raw PacBio reads using the CCS module in the SMRT Analysis software suite (v5.1, Pacific Biosciences). Subsequently, these CCSs were processed through the Iso-Seq pipeline (SMRT Link v6.0, Pacific Biosciences) for polishing, full-length transcript, and isoform clustering, resulting in a high-quality set of transcript isoforms. The resultant data were assembled into a non-redundant collection of Unigene transcripts specific to *P. pingi*. The raw sequencing data have been deposited in the National Genomics Data Center (<https://ngdc.cnpc.ac.cn/gsa/>) under accession number CRA016173.

The sequencing produced 554,382 CCS reads with an average length of 1,584 bp and an N50 of 1,674 bp. Following the removal of 35,421 chimeric and non-polyA full-length non-chimeric (FLNC) reads, a total of 499,007 polyA-containing FLNC reads were retained, exhibiting an average length of 1,412 bp and an N50 of 1,676 bp. These reads were clustered into 132,504 initial isoforms (average length: 1,673 bp; N50: 2,022 bp), which were further consolidated into 70,321 non-redundant isoforms (average length: 1,882 bp; N50: 2,243 bp). These high-quality isoforms served as the basis for subsequent analyses (Table S6).

### 2.2.3. TRANSCRIPTOME SEQUENCING

Transcriptome sequencing was performed on *P. pingi* individuals sourced from four distinct locations: the Yalong River Jinping–Guandi Fish Stock Enhancement and Release Station, the Heima Fish Stock Enhancement and Release



Station, Sichuan Runjie Hongda Aquatic Technology Co., Ltd., and Ya'an Zhougong River Yafish Co., Ltd.

For all samples, RNA extraction and sequencing were conducted using dorsal muscle tissue. Juvenile fish (average weight:  $9.86 \pm 0.8$  g) were anesthetized with MS-222 (100 mg/L), dissected under sterile conditions, and tissue samples were immediately placed in pre-labeled cryovials. The samples were flash-frozen in liquid nitrogen, and stored on dry ice until transportation to Beijing Biomarker Technologies Co., Ltd. for sequencing.

Total RNA extraction from muscle tissue was carried out using the TRIzol reagent (Invitrogen, Thermo Fisher Scientific, USA) method according to standard protocols. RNA concentration and purity were quantified using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA), while RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). RNA libraries were constructed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, USA) following the manufacturer's instructions. Specifically, poly-A mRNA was isolated via magnetic oligo(dT) beads, fragmented, and reverse-transcribed into cDNA. Double-stranded cDNA was subsequently amplified through PCR, and adapters were ligated for size selection to generate libraries compatible with the Illumina sequencing platform. Library quality was evaluated using a Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific, USA) and confirmed with an Agilent 2100 Bioanalyzer.

Sequencing was performed on the Illumina HiSeq 6000 platform (Illumina, San Diego, CA, USA) to generate 150 bp paired-end reads. The raw sequencing reads were subjected to quality assessment using FastQC (v0.11.9). Subsequently, adapter trimming and filtering of low-quality bases were carried out using Trimmomatic (v0.39). Clean reads were then aligned to the de novo assembled transcriptome of *P. pingi* using Trinity (v2.8.5) for transcript reconstruction. The raw data have been deposited in the National Genomics Data Center under accession number CRA024585.

## 2.3. DEVELOPMENT OF POLYMORPHIC SSR MARKERS

In summary, SSR-containing sequences were identified from the full-length transcriptome data, and potential polymorphic loci were predicted using SSREnricher software. Genomic DNA extracted from 12 parental individuals was utilized to validate 50 randomly selected primer pairs. Primers generating distinct and intense bands were preliminarily screened via 2% agarose gel electrophoresis, followed by further confirmation of polymorphism through 8% PAGE with silver staining. Subsequently, fluorescent capillary electrophoresis was employed to validate the polymorphism in selected SSR markers.

### 2.3.1. IDENTIFICATION AND CHARACTERIZATION OF SSR LOCI

To identify and characterize SSR loci of *P. pingi*, a comprehensive analysis was conducted on the Unigene sequences of full-length transcriptome data (CRA016173). The detection of SSR was performed using MISA (v2.1) software,

which facilitates the identification of six types of repeat motifs: mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats. The screening criteria were defined as follows: sequences exceeding 1000 bp in length were included, with a minimum of 10 repeats for mononucleotide motifs, 6 for dinucleotide motifs, and 5 for tri-, tetra-, penta-, and hexanucleotide motifs. Furthermore, the maximum allowable distance between two adjacent SSR loci in compound SSRs was set at 100 bp.

### 2.3.2. SCREENING OF POLYMORPHIC SSR MARKERS

To develop polymorphic simple sequence repeat (SSR) markers for *P. pingi*, full-length transcriptome data (CRA016173) and transcriptome sequencing data (CRA024585) were comprehensively analyzed to identify potential SSR loci. The detection of polymorphic SSR loci was performed using our previously developed tool, SSREnricher (v1.1).<sup>12</sup> SSREnricher (v1.1) enriches polymorphic SSRs by integrating SSR detection with transcriptome clustering information. It incorporates MISA (v2.1) for SSR identification and CD-HIT (v4.6.8) for sequence clustering. The software comprises six core analytical modules: SSR detection, sequence clustering, sequence modification, enrichment of polymorphic SSR-containing sequences, false positive removal, and result output with multiple sequence alignment. By comparing multiple transcriptome datasets, SSREnricher can predict loci with potential polymorphism, thereby significantly reducing the time and cost associated with experimental validation.

Following the identification of candidate SSR loci, primers were designed using Primer5 software according to standardized criteria: primer length ranging from 18 to 25 bp, GC content between 40% and 60%, annealing temperature within 48 – 62°C, avoidance of secondary structures, and an expected amplicon size of 200–350 bp for electrophoretic analysis. Primers containing three or more consecutive identical bases at the 3' end were excluded to minimize the risk of nonspecific amplification. All primers were synthesized by Beijing Qingke Biological Technology Co., Ltd (<https://tsingke.com.cn/>).

Genomic DNA was extracted using the Animal Tissue/Cell Genomic DNA Extraction Kit (D1700, Solarbio, Beijing, China). The concentration and quality of the DNA were evaluated using a spectrophotometer, and 2.0% agarose gel electrophoresis was performed to confirm single-band clarity. Subsequently, the DNA samples were diluted to a concentration of 50 ng/μL and stored at –20°C.

PCR amplification was carried out using the sanTaq PCR Mix reagent kit (Sogon). The PCR reaction mixture (total volume of 10 μL) contained 5 μL of 2× sanTaq PCR Mix, 1 μL genomic DNA, 3 μL ddH<sub>2</sub>O, and 0.5 μL each of forward and reverse primers (10 μmol/L). The thermocycling protocol was as follows: initial denaturation at 94°C for 5 min; 30 cycles of denaturation at 94°C for 30 s, annealing at 50–60°C for 30 s, and extension at 72°C for 30 s; followed by a final extension step at 72°C for 4 min. The PCR products were stored at 4°C and analyzed via 2.0% agarose gel electrophoresis to select primer pairs that produced bright and distinct bands.

### 2.3.3. VALIDATION OF SSR POLYMORPHISM

Further assessment of polymorphism was conducted using 8% non-denaturing polyacrylamide gel electrophoresis (PAGE), followed by silver staining. Electrophoresis was carried out at a voltage range of 60 – 80 V for 2 – 3 hours, contingent upon the migration distance. The resulting bands were visualized and documented using digital imaging techniques.

Nine polymorphic SSR loci were randomly chosen for validation using fluorescently labeled primers. Forward primers were labeled at the 5' end with FAM, HEX, or ROX fluorophores. The PCR products were preliminarily validated via 2% agarose gel electrophoresis and subsequently purified for capillary electrophoresis, which was performed by Beijing Qingke Biotechnology Co., Ltd. Genotyping was conducted using fluorescence-based detection and analyzed with GeneMarker (v1.75; <https://softgenetics.com/products/genemarker/>). Given that *P. pingi* is a tetraploid species ( $4n = 98$ ), manual curation was essential to ensure accurate genotype calling. Since automatic allele calling by the software may only record one or two alleles based on peak numbers, manual interpretation of allele counts and peak intensity ratios was implemented to refine genotyping (Fig. S1).

### 2.3.4. GENETIC DIVERSITY ANALYSIS OF SSR LOCI

Genetic diversity analyses were conducted on the validated SSR loci. The Shannon diversity index (I) was calculated using ATetra, whereas other parameters were calculated using GenoDive (v3.0), which supports polyploid data. The metrics analyzed included the number of alleles (Na), allele frequency (AF), the effective number of alleles (Ne), expected heterozygosity (He), observed heterozygosities (Ho), and Hardy-Weinberg equilibrium (HWE). Furthermore, polymorphism information content (PIC) and observed heterozygosity (Ho) were computed using standard formulas.

PIC was calculated as follows:

$$PIC = 1 - \sum_{i=1}^n p_i^2 - 2 \left[ \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i^2 p_j^2 \right] \quad (1)$$

Where  $n$  is the total number of observed alleles at the locus, and  $p_i$  and  $p_j$  represent the frequencies of the  $i_{th}$  and  $j_{th}$  alleles, respectively. According to the classification criteria, an SSR locus is considered highly polymorphic when  $PIC > 0.5$ , moderately polymorphic when  $0.25 < PIC < 0.5$ , and lowly polymorphic when  $PIC < 0.25$ .

Observed heterozygosity (Ho) was calculated as:

$$Ho = \frac{\text{Number of observed heterozygous individuals}}{\text{Total number of individuals}}$$

### 2.4. PARENTAGE IDENTIFICATION

The validation of the nine SSR loci was conducted using 300 offspring from single-pair families and 270 offspring from the mixed-rearing group, following the procedures outlined in Section 2.4.3. The genotype data of the 12 parents and 570 offspring were then combined to calculate the genetic

diversity parameters of the parental and offspring populations, including I, Na, AF, Ne, He, HWE, PIC, and Ho.

To assess the accuracy of parentage assignment, we conducted computer simulations using the FaMoz software.<sup>13</sup> The cumulative probability of exclusion (CPE) was calculated based on allele frequency data for each SSR locus (Table S7), with markers treated as codominant. The exclusion probability (EP) for individual SSR loci and the CPE across multiple loci were computed using the “Exclusion Probabilities” module under the “Probabilities” submenu. Parentage assignment was subsequently performed using both Cervus (v3.0) and PAPA (v2.0), which infer parent-offspring relationships based on exclusion principles. All 300 single-pair family offspring and 270 offspring from the mixed-rearing rearing group were analyzed with both programs to ensure robust identification under varying rearing conditions. In Cervus, simulations were conducted under the assumption that both maternal and paternal genotypes were known, with a confidence level set at 95%. In PAPA, analysis was performed using a uniform error model with a total error rate of 0.02. Offspring were assigned to their most likely parents only when consistent matches were identified by both Cervus and PAPA; these concordant results were considered true parent-offspring pairs.

### 2.5. ANALYSIS OF GENETIC DISTANCE AMONG PARENTAL INDIVIDUALS.

To evaluate the genetic distances among parental individuals, we employed the R packages poppr and adegenet. The raw SSR genotype data, comprising nine loci with four alleles per individual, were initially formatted by combining alleles at each locus into a hyphen-separated string (e.g., “16/26/28/30”). Subsequently, these formatted genotypes were transformed into a genind object using the df2genind function in adegenet, with ploidy explicitly set to four to account for the tetraploid nature of the individuals.

We applied the recode\_polyploids function in poppr to standardize polyploid data prior to distance computation. Bruvo's genetic distance was then calculated using the bruvo.dist function, specifying the repeat length of each locus based on the SSR motif (di-, tri-, or tetra-nucleotide repeats). Bruvo's genetic distance ranges from 0 (indicating high genetic similarity) to 1 (indicating greater genetic divergence).

Correlation analyses were performed using the ggpubr package, and linear regression equations were generated using ggpmisc. All visualizations were created with ggplot2.

### 2.6. ESTIMATION OF GENETIC PARAMETERS FOR GROWTH TRAITS

A total of 270 individuals from the mixed-rearing groups of *P. pingi* were measured for three growth traits: TL, BL, and BW. For detailed analysis, parentage identification results and corresponding growth trait data of each family were compiled into an Excel spreadsheet. Normality tests of all offspring growth data were performed using the Kolmogorov-Smirnov (K-S) test in SPSS (v.23.0). Basic descriptive statistics, including mean, standard deviation (SD),

maximum, and minimum values, were computed. Additionally, the coefficient of variation (CV) was calculated in Excel using the following formula:

$$CV = \frac{SD}{Mean} * 100\% \quad (3)$$

Genetic parameters and breeding values (BLUPs) were estimated using ASReml 3.0 software (<https://vsni.co.uk/software/asreml-r/>). The estimation was conducted via an animal model, utilizing the restricted maximum likelihood (REML) method for variance component estimation and the best linear unbiased prediction (BLUP) approach for breeding value estimation.

The animal model applied in this study is as follows:

$$Y_{ijk} = \mu + f_{ijk} + a_{ijk} + e_{ijk} \quad (4)$$

Where  $i$ ,  $j$ , and  $k$  represent the sire, dam, and individual ID, respectively.  $Y_{ijk}$  is the observed phenotypic value of the trait;  $\mu$  is the overall population mean;  $f_{ijk}$  represents the family effect (including common environmental effects);  $a_{ijk}$  is the additive genetic effect; and  $e_{ijk}$  is the random residual. Among these,  $\mu$  and  $f_{ijk}$  were treated as fixed effects, while the others were treated as random effects. All random effects were assumed to follow a normal distribution and to be mutually independent.

Narrow-sense heritability ( $h^2$ ) was estimated using the `nvpredict` function in ASReml 3.0, based on the following formula:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (5)$$

Where  $h^2$  is heritability,  $\sigma_a^2$  is the additive genetic variance, and  $\sigma_e^2$  is the residual (phenotypic) variance. In general, traits with  $h^2$  values less than 0.1 are considered to have low heritability, those with  $0.1 \leq h^2 \leq 0.3$  exhibit moderate heritability, and values greater than 0.3 indicate high heritability.<sup>14</sup>

Genetic and phenotypic correlations between traits were estimated, using standard covariance-based formulas in ASReml 3.0:

$$r_G = \frac{COV_{G_{XY}}}{\sqrt{\sigma_{G_X}^2 \sigma_{G_Y}^2}} \quad (6)$$

$$r_P = \frac{COV_{P_{XY}}}{\sqrt{\sigma_{P_X}^2 \sigma_{P_Y}^2}} \quad (7)$$

Where  $r_G$  and  $r_P$  represent the genetic and phenotypic correlations between traits X and Y, respectively.  $COV_{G_{XY}}$  and  $COV_{P_{XY}}$  are the genetic and phenotypic covariance components between the two traits, and  $\sigma_G^2$ ,  $\sigma_P^2$  are the corresponding variances.

## RESULTS

### 3.1. SCREENING OF POLYMORPHIC MICROSATELLITES

A total of 30,147 sequences harboring SSR loci were detected in the full-length transcriptome data. Using SSRENricher software, 548 potentially polymorphic SSRs were identified. From these, 50 SSR loci were randomly selected for polymorphism assessment. Among them, 22 primer pairs produced clear and well-defined amplification bands

(Table S8). Subsequent analysis using non-denaturing PAGE confirmed polymorphism at 14 of these loci, evaluating the level of polymorphism of these primers in parents. Ultimately, nine highly specific and efficient SSR markers were validated using fluorescent capillary electrophoresis (Table 1).

All nine SSR loci demonstrated high levels of polymorphism among the 12 parental individuals, with PIC values exceeding 0.5 (Table S9).

### 3.2. PARENTAGE IDENTIFICATION SYSTEM

Genetic diversity analysis of the *P. pingi* parental individuals and their  $F_1$  offspring revealed a total of 65 alleles across the 12 parents and 570 offspring at the nine SSR loci. The mean  $N_a$  was 7.222, while the average  $N_e$  was 3.503. The mean  $H_o$  was 0.936, while the  $H_e$  averaged 0.692. The average PIC was 0.647, with all loci—except PP021 and PP001—showing PIC values above 0.5, indicating high levels of polymorphism. The mean  $I$  was 1.407. HWE analysis indicated that only loci PP021 and PP047 conformed to HWE expectations; all other loci showed significant deviations (Table 2).

Based on simulations conducted using FaMoz, the results indicated that in the absence of known parental genotypes, the exclusion probability per locus (E-PP) ranged from 0.480 to 0.771, with an average of 0.696, reflecting strong power to exclude non-parental assignments (Table S10). The cumulative probability of exclusion (CPE-PP) across all nine SSR markers reached 0.999984, which exceeds the recommended threshold of 0.9999 for reliable parentage identification (Fig. 2).

Among the 300 offspring from single-pair families, 288 individuals (96%) were assigned to the same parental pairs by both Cervus and PAPA, confirming the high accuracy of the combined parentage identification approach. Based on the established parentage identification system, the results from the mixed-rearing groups indicated that the dam ♀2 exhibited higher reproductive efficiency when crossed with these sires (Fig. S2).

### 3.3. RELATIONSHIP BETWEEN PARENTAL GENETIC DISTANCE AND OFFSPRING GROWTH

Bruvo's genetic distances among parental individuals ranged from 0.33 to 0.40, with a mean of 0.37 (Table S11). Based on parentage assignments derived from the mixed-rearing group, we estimated the growth rates of offspring from different families. A linear regression analysis conducted between parental genetic distance and offspring growth rate resulted in the equation:  $Y = 4.16 + 4.24X$  ( $R = 0.089$ ,  $P = 0.15$ ) (Fig. 3), which indicates no statistically significant relationship between the two variables.

### 3.4. GENETIC ANALYSIS OF GROWTH TRAITS

K-S tests confirmed that all three traits followed a normal distribution, meeting the assumptions required for subsequent statistical analyses. The CV indicated differing levels of phenotypic variation among the traits. Body weight

**Table 1. Basic information of nine polymorphic SSR Loci in *Percocypris pingi*.**

| Locus | Repeat motif | Annealing temperatures (°C) | Segment length (bp) | Primer sequence(5'~3')                                   | Labeled dye |
|-------|--------------|-----------------------------|---------------------|----------------------------------------------------------|-------------|
| PP021 | (AC/GT)6     | 55.0                        | 300-318             | F: TATGATGCAATCAATGAATGG<br>R: CGAAACATCTGGACAGAAG       | 5'FAM       |
| PP001 | (CA/TG)12    | 54.0                        | 270-298             | F: GAATCCCTTACAACCTTACC<br>R: CGTTAAAGGAATGTCAGAAT       | 5'HEX       |
| PP007 | (AC/GT)10    | 54.0                        | 250-264             | F: GTCTTCTGTAATCTGTAACCTGC<br>R: TGAGTAGCACATAATCTTAAGCT | 5'ROX       |
| PP028 | (CA/TG)11    | 55.3                        | 270-278             | F: GTGCTACCTGTGCCATCC<br>R: GGAGGGTTAGTATGATTACACG       | 5'FAM       |
| PP034 | (AC/GT)11    | 58.1                        | 268-278             | F: AAGATTCCTTTCCGTCAAGAC<br>R: GTCGTCGGTGTAAGTGTCATC     | 5'HEX       |
| PP016 | (AC/GT)17    | 60.0                        | 318-328             | F: AAGTGAAGAGCGGAGTTG<br>R: AATCTCATCGTCTGTCTG           | 5'ROX       |
| PP006 | (AC/GT)12    | 54.0                        | 268-300             | F: GAAAGCCCTAGAAAATATTGCT<br>R: TGGGTTTTACTGACCAGAGG     | 5'FAM       |
| PP047 | (AC/GT)21    | 55.2                        | 258-284             | F: GAAGAGATGGAGGTGGATGT<br>R: CTGATTGGTCAAGGAGGGAT       | 5'HEX       |
| PP042 | (CA/TG)11    | 55.2                        | 282-290             | F: GTTCCAGACGCTCCCTA<br>R: TTGCGCCTCAGTGTATACT           | 5'ROX       |

**Table 2. Genetic diversity of nine SSR loci in the parent and offspring populations of *Percocypris pingi*<sup>a</sup>**

| Locus | Na    | Ne    | Ho    | He    | PIC   | I     | HWE |
|-------|-------|-------|-------|-------|-------|-------|-----|
| PP021 | 8     | 2.082 | 0.789 | 0.520 | 0.475 | 1.111 | ns  |
| PP001 | 7     | 2.177 | 0.861 | 0.541 | 0.473 | 1.041 | **  |
| PP007 | 8     | 3.032 | 0.811 | 0.671 | 0.675 | 1.470 | **  |
| PP028 | 5     | 4.004 | 0.997 | 0.751 | 0.707 | 1.477 | **  |
| PP034 | 5     | 3.794 | 1.000 | 0.737 | 0.675 | 1.414 | **  |
| PP016 | 6     | 3.306 | 0.990 | 0.698 | 0.637 | 1.364 | **  |
| PP006 | 10    | 3.705 | 0.997 | 0.730 | 0.675 | 1.455 | **  |
| PP047 | 10    | 4.914 | 0.988 | 0.797 | 0.771 | 1.793 | ns  |
| PP042 | 6     | 4.512 | 0.997 | 0.779 | 0.736 | 1.542 | **  |
| AVE   | 7.222 | 3.503 | 0.936 | 0.692 | 0.647 | 1.407 | -   |

<sup>a</sup> Note: "ns" indicates compliance with Hardy-Weinberg equilibrium, and "\*" indicates a significant deviation from Hardy-Weinberg equilibrium after Bonferroni correction ( $P < 0.05$ ), "\*\*\*" indicates a highly significant deviation ( $P < 0.01$ )

exhibited the highest variation (CV = 24.41%), significantly greater than that of total length (CV = 9.71%) and body length (CV = 10.51%) (Table S12).

Heritability estimates based on univariate animal models are presented in Table 3. The heritability of TL was  $0.488 \pm 0.119$ , indicating that 48.8% of the phenotypic variance in TL is attributable to additive genetic effects. BL had a heritability of  $0.234 \pm 0.072$ , while BW showed a heritability of  $0.219 \pm 0.069$ . According to standard classification criteria, TL is considered a trait with high heritability, whereas BL and BW fall into the moderate heritability category. In the multivariate animal model (Table 4), heritability estimates changed notably. The heritability of TL decreased to  $0.171 \pm 0.033$ , while body length remained similar at  $0.210 \pm 0.050$ . Interestingly, the heritability of body weight in-

creased to  $0.307 \pm 0.082$ . Under the multivariate animal model, the heritability of BW is categorized as a high-heritability trait. In contrast, the heritability estimates for TL and BL remained within the moderate heritability range.

Genetic and phenotypic correlations among TL, BL, and BW were analyzed (Table S13). At the genetic level, the highest correlation was observed between TL and BL ( $0.976 \pm 0.003$ ), followed by TL and BW ( $0.935 \pm 0.009$ ), and BL and BW ( $0.933 \pm 0.011$ ). At the phenotypic level, TL and BL exhibited the strongest correlation ( $0.988 \pm 0.016$ ), followed by TL and BW ( $0.965 \pm 0.050$ ), and BL and BW ( $0.921 \pm 0.080$ ). All pairwise correlations were statistically significant ( $P < 0.05$ ).



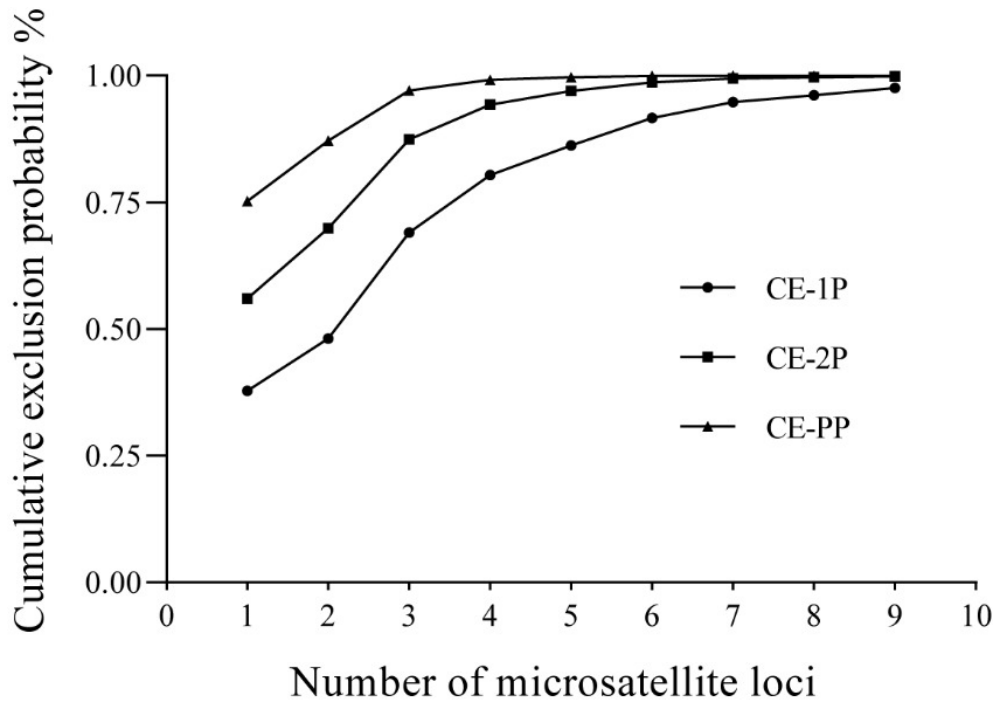


Fig. 2. Cumulative exclusion probability across nine SSR loci in *Percocypris pingi*. The x-axis represents the number of microsatellite loci, and the y-axis shows the cumulative exclusion probability for different parentage testing categories, including one-parent (CE-1P), two-parent (CE-2P), and parent-pair (CE-PP) exclusion probabilities.

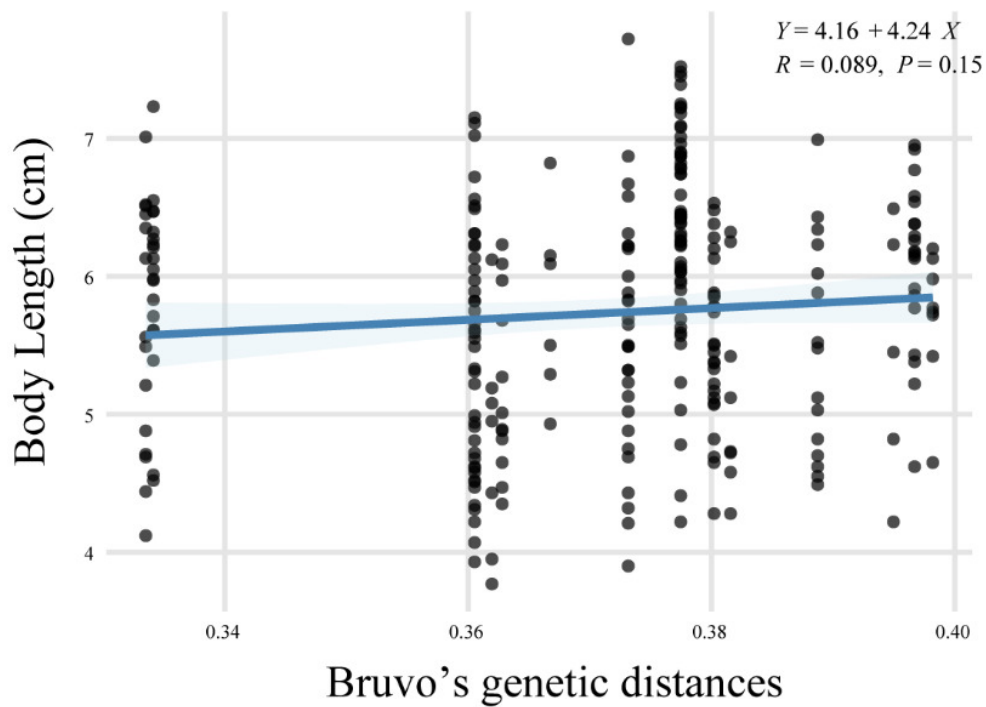


Fig. 3. Relationship between parental genetic distance and offspring growth. The scatter plot illustrates the regression between parental genetic distance (x-axis, Bruvo's genetic distance) and offspring body length (y-axis).



**Table 3. Heritability estimation based on a single trait animal model.<sup>a</sup>**

| Trait           | $\sigma_a^2$ | $\sigma_e^2$ | $h^2 \pm SE$      |
|-----------------|--------------|--------------|-------------------|
| Total length/cm | 0.480        | 0.51         | 0.488 $\pm$ 0.119 |
| Body length/cm  | 0.220        | 0.70         | 0.234 $\pm$ 0.072 |
| Body weight/g   | 1.255        | 5.465        | 0.219 $\pm$ 0.069 |

<sup>a</sup> Note:  $h^2$  represents heritability; SE stands for Standard Error

**Table 4. Heritability estimation based on multi trait animal model.<sup>a</sup>**

| Trait           | $\sigma_a^2$ | $\sigma_e^2$ | $h^2 \pm SE$      |
|-----------------|--------------|--------------|-------------------|
| Total length/cm | 0.200        | 0.970        | 0.171 $\pm$ 0.033 |
| Body length/cm  | 0.190        | 0.720        | 0.210 $\pm$ 0.050 |
| Body weight/g   | 1.927        | 5.343        | 0.307 $\pm$ 0.082 |

<sup>a</sup> Note:  $h^2$  represents heritability; SE stands for Standard Error

## DISCUSSION

### 4.1. METHODS FOR THE IDENTIFICATION OF SSR LOCI

Although the genomes of numerous economically significant aquaculture species have been sequenced, genomic resources for many endangered fish species remain limited. Consequently, transcriptome sequencing has emerged as a primary method for generating extensive sequence information and developing polymorphic SSR markers in these species. Third-generation full-length transcriptome sequencing provides the advantage of long read lengths, enabling the direct acquisition of complete mRNA sequences and structural information without the need for cDNA fragmentation or assembly. This approach has become a critical strategy for developing polymorphic SSR markers based on transcriptome data.<sup>15</sup> In our study, a total of 58,021 unigenes were obtained, and SSR loci were identified in 30,147 unigenes, resulting in an SSR occurrence frequency of 51.96%. This frequency is significantly higher than that reported in many other fish species, such as Patagonian toothfish *Dissostichus eleginoides* (Smitt, 1898) (30.46%),<sup>16</sup> indicating that third-generation transcriptome sequencing is a more effective approach than traditional transcriptome sequencing for the development of SSR markers. In this study, we employed the SSREnricher software to mine polymorphic loci across *P. pingi* populations from multiple geographic locations. As a result, the polymorphic SSR loci identified here may exhibit broader applicability. This is particularly important because previous studies have indicated that parentage identification systems developed from samples of limited geographic origin may compromise accuracy.<sup>7</sup>

### 4.2. SCREENING OF POLYMORPHIC SSR LOCI IN *P. PINGI*

Among the 50 designed SSR primer pairs, 14 were successfully amplified and exhibited clear polymorphism. By integrating second- and third-generation sequencing data using SSREnricher, we identified that 63.64% of the 22 se-

lected clear bands contained polymorphic SSRs. According to Luo et al.,<sup>12</sup> over 90% of markers predicted by SSREnricher as polymorphic are indeed polymorphic in practice. However, it is evident that the probability of detecting polymorphic SSRs decreases in polyploid species. A similar phenomenon was reported in loach *Misgurnus anguillicaudatus* (Cantor, 1842), where only 24.44% of loci that were polymorphic in diploids remained highly polymorphic in tetraploids.<sup>17</sup> One possible explanation is that, in polyploids, some alleles may fail to amplify due to mutations at primer-binding sites, leading to an underestimation of allele numbers.

A more critical consideration in the screening of polymorphic SSRs is determining the minimum number of loci required to construct a reliable parentage identification system. Based on our literature review, most studies suggest that at least seven polymorphic loci are necessary to achieve sufficient accuracy for parentage assignment.<sup>18,19</sup> Notably, Wang et al.<sup>19</sup> demonstrated that a set of 12 polymorphic loci could support the establishment of two independent parentage identification systems. Therefore, selecting at least eight highly polymorphic loci is considered a prudent and effective strategy. When more than 12 loci are available, constructing multiple complementary parentage identification panels becomes feasible.

### 4.3. ESTABLISHMENT OF A PARENTAGE ASSIGNMENT SYSTEM

To establish a reliable parentage identification system, we first evaluated key indicators of population genetic diversity, including Na, Ne, I, Ho, He, and PIC. A total of 65 alleles were detected across both the parent and offspring populations, with an average of 7.222 alleles per locus and an average Ne of 3.503. The mean Ho was 0.936, and the He was 0.692, indicating a high level of genetic diversity in the population. The average PIC value was 0.647, with all except PP021 (0.475) and PP001 (0.473)-exceeding 0.5, demonstrating that the majority of loci were highly polymorphic. The average I was 1.407, further supporting the

conclusion of substantial genetic diversity. Overall, the genetic diversity parameters of the nine SSR loci used in this study indicate that *P. pingi* populations harbor sufficient variation, making these markers suitable for constructing a robust parentage identification system.

The PIC is a critical parameter for evaluating the effectiveness of parentage identification systems. Typically, SSR loci with a PIC value greater than 0.5 suggest the presence of multiple alleles with relatively even frequency distributions, which is advantageous for individual identification and accurate parentage assignment. In this study, two loci exhibited PIC values slightly below 0.5, specifically 0.475 and 0.473. However, a previous study suggested that loci with PIC values above 0.4 may still be informative and useful for parentage analysis.<sup>20</sup> Subsequent validation confirmed that these two loci were indeed effective in the parentage identification system. Nonetheless, we do not recommend the inclusion of loci with PIC values lower than 0.5, as the initial validation involving 12 parental individuals demonstrated that all selected loci exhibited PIC values exceeding 0.5. Surprisingly, our study demonstrated that the majority of microsatellite loci adhered to HWE. Napora-Rutkowski et al.<sup>21</sup> noted that deviations from HWE are relatively common in captive fish populations, primarily attributed to hatchery practices involving a limited number of founder individuals and unbalanced sex ratios. A comparable observation was made by Sánchez-Velásquez et al.,<sup>5</sup> where eight out of 10 polymorphic loci showed deviations from HWE. However, in their study, the influence of inbreeding was excluded due to the use of wild populations, and a high frequency of null alleles (Null Freq) was identified as a significant contributing factor. In contrast, inbreeding is likely a critical factor contributing to the observed deviations in our study. It is also essential to evaluate whether deviations from HWE could compromise the accuracy of the parentage identification system. Theoretically, HWE deviations may alter allele frequency estimates, thereby reducing the statistical confidence of parentage assignment. Nevertheless, in systems utilizing multiple polymorphic markers, the impact of HWE deviations appears to be negligible.<sup>5</sup> Consequently, rather than focusing exclusively on HWE status, greater emphasis should be placed to the PIC of each locus. Extensive research has highlighted that low PIC values can elevate the risk of misassignment and increase analytical cost, whereas the influence of HWE deviations on parentage analysis remains generally limited.<sup>5,20</sup>

Using nine highly polymorphic SSR loci, we conducted parentage analysis on 12 candidate parents and 300 offspring individuals from single-pair families. The combined CPE reached 0.999984, corresponding to a simulated accuracy of 99.99%, which significantly exceeds the threshold for confirmed parentage ( $\geq 99.73\%$ ). In this study, the actual assignment accuracy for single-family *P. pingi* offspring was 96%, slightly lower than the simulated value but still sufficiently high for practical applications. In other studies, parentage accuracy based on 12 highly polymorphic loci reached 100% in mixed-family settings, with an accuracy of 98.69% in the test population. In grass carp *Ctenopharyngodon*

*idella* (Valenciennes, 1844), 12 microsatellite loci were optimized into three multiplex PCR systems for parentage identification, and 99.6% of offspring were successfully assigned to a specific parental pair.<sup>22</sup> These findings demonstrate the feasibility of using diploid analysis software, such as Cervus v3.0 and PAPA, to infer parentage even in polyploid fish species.

#### 4.4. GENETIC DISTANCE ANALYSIS AND ITS RELATIONSHIP WITH OFFSPRING GROWTH

Regulating the genetic distance between breeding pairs can significantly enhance offspring survival and average BL.<sup>3</sup> A low genetic distance increases the likelihood of inbreeding depression, which can lead to reduced fitness and viability of the offspring. A study demonstrated that wild populations with greater genetic distances exhibited significantly higher growth rates compared to cultured populations with smaller genetic distances.<sup>23</sup> This highlights the importance of maintaining an effective breeding population by selecting parent pairs with moderate to high genetic divergence. In this study, no definitive results were obtained, as the positive correlation between growth traits and genetic distance was not statistically significant. This may primarily result from the relatively narrow genetic background of the broodstock, which limited the overall range of genetic divergence and reduced the statistical power to detect associations. In addition, the sample size of breeding pairs was relatively small, which may have weakened the robustness of the correlation analysis. Moreover, the relationship between genetic distance and heterosis is likely non-linear, with both insufficient and excessive genetic divergence potentially leading to suboptimal offspring growth. Non-additive genetic effects and environmental factors may also obscure the true genetic relationship between parental distance and offspring growth.

#### 4.5. HERITABILITY AND BREEDING VALUE EVALUATION

In genetic breeding of aquatic animals, the heritability of growth traits serves as a critical indicator for evaluating the degree to which phenotypic variation is influenced by genetic factors. A higher heritability suggests a greater genetic contribution to trait variation, thereby indicating stronger potential for genetic improvement via selective breeding. In this study, the estimated heritability of growth traits in *P. pingi* exhibited differences between the univariate and multivariate models. Specifically, under the univariate model, the heritability estimates were  $0.488 \pm 0.119$  for TL,  $0.234 \pm 0.072$  for BL, and  $0.219 \pm 0.069$  for BW. Conversely, under the multivariate model, the heritability estimates were  $0.171 \pm 0.033$  for TL,  $0.210 \pm 0.050$  for BL, and  $0.307 \pm 0.082$  for BW. Similarly, previous studies on Asian seabass *Lates calcarifer* (Bloch, 1790) reported heritability estimates ranging from 0.11 – 0.32 for BL and 0.12 – 0.34 for BW, which align within a comparable range.<sup>24</sup> These findings underscore significant discrepancies in heritability estimates for TL and BL between the two models, while the heritability of BW increased in the multivariate model. Such inconsistencies may arise due to genetic corre-

lations among traits. The multivariate model incorporates covariance structures between traits, potentially enhancing the accuracy of heritability estimates, particularly for traits with low individual heritability or strong genetic correlations. Further analysis of both genetic and phenotypic correlations among TL, BL, and BW revealed values exceeding 0.9 ( $P < 0.05$ ), indicating robust intertrait relationships. Consequently, the multivariate approach is deemed more suitable for this study. Notably, BW emerged as a trait of high breeding relevance and could serve as a key target in future selection programs. In general, traits with high heritability can be effectively improved through mass selection, whereas those with low heritability necessitate family-based selection strategies to achieve enhanced breeding efficiency.

In practical terms, the developed parentage identification system enables mixed-family communal rearing with reliable pedigree reconstruction (>95–99% assignment), which is essential for estimating individual EBVs under BLUP/GBLUP, implementing factorial matings, and running selection in commercial settings. By linking growth phenotypes to reconstructed pedigrees, breeders can rank broodstock on EBVs for BW (the most promising target in this study) while constraining coancestry to control inbreeding and maintain effective population size across generations. Nevertheless, the current work is limited by the moderate sample size and potential environmental influences on growth traits. Future studies integrating larger breeding populations, multi-generation validation, and genomic datasets will further enhance the accuracy and applicability of the developed system.

## ACKNOWLEDGMENTS

Individuals other than authors can be acknowledged by a brief statement immediately following the text. Acknowledgments of grant support should also be placed here.

## AUTHORS' CONTRIBUTION

Conceptualization: Siyi Zhu (Equal), Wei Luo (Equal), Zongjun Du (Equal). Formal Analysis: Siyi Zhu (Equal), Xing Ying (Equal), Wei Luo (Equal). Investigation: Siyi Zhu (Equal), Longjun Deng (Equal), Tiancai Li (Equal), Tong Li (Equal), Qingjie Wang (Equal), Shun Zhong (Equal), Xing Ying (Equal), Fulin Dai (Equal), Dongjie Wang (Equal). Validation: Siyi Zhu (Equal). Writing – original draft: Siyi Zhu

(Equal), Wei Luo (Equal). Methodology: Dongjie Wang (Equal), Wei Luo (Equal), Zongjun Du (Equal). Funding acquisition: Wei Luo (Equal), Zongjun Du (Equal). Project administration: Wei Luo (Equal), Zongjun Du (Equal). Writing – review & editing: Wei Luo (Equal), Zongjun Du (Equal). Data curation: Zongjun Du (Lead).

## COMPETING OF INTEREST – COPE

The authors declare there are no competing interests.

## ETHICAL CONDUCT APPROVAL – IACUC

All experimental protocols in this study were approved by the Animal Research and Ethics Committee of Sichuan Agricultural University (SAU20230310). Fish treatments were conducted in accordance with the guidelines for the Care and Use of Agricultural Animals in Research and Teaching, authorized by the Science and Technology Bureau of China. Approval from the Department of Wildlife Administration was not required for the experiments described in this paper.

## INFORMED CONSENT STATEMENT

All authors and institutions have confirmed this manuscript for publication.

## DATA AVAILABILITY STATEMENT

The transcriptome sequence data supporting the findings of this study are publicly available in the National Genomics Data Center (<https://ngdc.cncb.ac.cn/gsa/>) under accession numbers CRA016173 and CRA024585.

## FUNDING

This work was supported by the National Natural Science Foundation of China (No. 32172998), the Yalong River *Percocypris pingi* Germplasm Characterization and Its Application in Artificial Propagation Project (No. 000023-22XB0141), the National Modern Agricultural Industry Technology System – Sichuan Innovation Team (No. SCCXTD-2025-15), and the Study and Demonstration on the Breeding of New Cobitidae Varieties and Standardized Aquaculture Techniques Project (No. 2021YFYZ0015).

Submitted: September 18, 2025 CDT. Accepted: October 17, 2025 CDT. Published: December 30, 2025 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## REFERENCES

1. Grant WS, Jasper J, Bekkevold D, Adkison M. Responsible genetic approach to stock restoration, sea ranching and stock enhancement of marine fishes and invertebrates. *Rev Fish Biol Fisheries*. 2017;27:615-649. doi:[10.1007/s11160-017-9489-7](https://doi.org/10.1007/s11160-017-9489-7)
2. O'Sullivan RJ, Aykanat T, Johnston SE, et al. Captive-bred Atlantic salmon released into the wild have fewer offspring than wild-bred fish and decrease population productivity. *Proc R Soc B*. 2020;287(1937):20201671. doi:[10.1098/rspb.2020.1671](https://doi.org/10.1098/rspb.2020.1671)
3. Shan B, Liu Y, Song N, et al. Parentage determination of black sea bream (*Acanthopagrus schlegelii*) for stock enhancement: Effectiveness and loss of genetic variation. *Acta Oceanol Sin*. 2021;40:41-49. doi:[10.1007/s13131-020-1697-7](https://doi.org/10.1007/s13131-020-1697-7)
4. Wellmann R, Bennewitz J. Key genetic parameters for population management. *Front Genet*. 2019;10:667. doi:[10.3389/fgene.2019.00667](https://doi.org/10.3389/fgene.2019.00667)
5. Sánchez-Velásquez JJ, Pinedo-Bernal PN, Reyes-Flores LE, Yzásiga-Barrera C, Zelada-Mázmela E. Genetic diversity and relatedness inferred from microsatellite loci as a tool for broodstock management of fine flounder *Paralichthys adspersus*. *Aquaculture and Fisheries*. 2022;7(6):664-674. doi:[10.1016/j.aaf.2021.06.008](https://doi.org/10.1016/j.aaf.2021.06.008)
6. Luo W, Zhang N, Wang D, et al. Assessment of the heterosis of F1 hybrids of *Misgurnus anguillicaudatus* and *Paramisgurnus dabryanus* using microsatellite-based parentage assignment. *Aquac Int*. 2021;29:1331-1341. doi:[10.1007/s10499-021-00700-x](https://doi.org/10.1007/s10499-021-00700-x)
7. Jones AG, Small CM, Paczolt KA, Ratterman NL. A practical guide to methods of parentage analysis. *Mol Ecol Resour*. 2010;10(1):6-30. doi:[10.1111/j.1755-0998.2009.02778.x](https://doi.org/10.1111/j.1755-0998.2009.02778.x)
8. Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR, Raadsma HW. Genomic selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front Genet*. 2019;9:693. doi:[10.3389/fgene.2018.00693](https://doi.org/10.3389/fgene.2018.00693)
9. Li X, Deng Y, Yang K, et al. Genetic diversity and structure analysis of *Percocypris pingi* (Cypriniformes: Cyprinidae): implications for conservation and hatchery release in the Yalong River. *PLoS One*. 2016;11(12):e0166769. doi:[10.1371/journal.pone.0166769](https://doi.org/10.1371/journal.pone.0166769)
10. Deng Y, Yang K, Gan W, Zeng R, Yan P, Song Z. Development of 12 tetranucleotide microsatellite markers for the tetraploidy fish *Percocypris pingi* (Tchang). *Conserv Genet Resour*. 2015;7:99-101. doi:[10.1007/s12686-014-0302-9](https://doi.org/10.1007/s12686-014-0302-9)
11. Wu S, Zeng R, Wang H. Application of Tagging Techniques in Fish Stock Enhancement in the Yalong River. *Yangtze River*. 2020;51(S1):56-60. doi:[10.16232/j.cnki.1001-4179.2020.S1.014](https://doi.org/10.16232/j.cnki.1001-4179.2020.S1.014)
12. Luo W, Wu Q, Yang L, et al. SSREnricher: a computational approach for large-scale identification of polymorphic microsatellites based on comparative transcriptome analysis. *PeerJ*. 2020;8:e9372. doi:[10.7717/peerj.9372](https://doi.org/10.7717/peerj.9372)
13. Gerber S, Chabrier P, Kremer A. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Mol Ecol Notes*. 2003;3(3):479-481. doi:[10.1046/j.1471-8286.2003.00439.x](https://doi.org/10.1046/j.1471-8286.2003.00439.x)
14. Falconer DS. *Introduction to Quantitative Genetics*. Pearson Education India; 1996:224-228. doi:[10.1007/978-3-030-83940-6\\_1](https://doi.org/10.1007/978-3-030-83940-6_1)
15. Zhang J, Liu J, Han C, et al. Full-length transcriptome sequence and SSR marker development for genetic diversity research in yellowfin seabream *Acanthopagrus latus*. *J Oceanol Limnol*. 2023;41(3):1073-1083. doi:[10.1007/s00343-022-1311-2](https://doi.org/10.1007/s00343-022-1311-2)
16. Touma J, García KK, Bravo S, et al. De novo assembly and characterization of Patagonian toothfish transcriptome and develop of EST-SSR markers for population genetics. *Front Mar Sci*. 2019;6:720. doi:[10.3389/fmars.2019.00720](https://doi.org/10.3389/fmars.2019.00720)
17. Feng B, Yi SV, Zhang M, Zhou X. Development of novel EST-SSR markers for ploidy identification based on de novo transcriptome assembly for *Misgurnus anguillicaudatus*. *PLoS One*. 2018;13(4):e0195829. doi:[10.1371/journal.pone.0195829](https://doi.org/10.1371/journal.pone.0195829)
18. Dong JY, Huang YX, Chen Z, Xu XW, Zhang XM. Embryonic development and microsatellite-based parentage assignment of seaweed pipefish *Syngnathus schlegelii*. *Hydrobiologia*. 2020;847:913-923. doi:[10.1007/s10750-019-04152-6](https://doi.org/10.1007/s10750-019-04152-6)
19. Wang X, Hua Q, Wu L, Weng Z, Huang W, Meng Z. Development of a multiplex microsatellite assay for parentage assignment in orange-spotted groupers (*Epinephelus coioides*). *J World Aquac Soc*. 2021;52(2):435-444. doi:[10.1111/jwas.12774](https://doi.org/10.1111/jwas.12774)



20. Tibihika PD, Curto M, Dornstauder-Schrammel E, et al. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conserv Genet.* 2019;20:357-372. doi:[10.1007/s10592-018-1136-x](https://doi.org/10.1007/s10592-018-1136-x)
21. Napora-Rutkowski Ł, Rakus K, Nowak Z, et al. Genetic diversity of common carp (*Cyprinus carpio* L.) strains breed in Poland based on microsatellite, AFLP, and mtDNA genotype data. *Aquaculture.* 2017;473:433-442. doi:[10.1016/j.aquaculture.2017.03.005](https://doi.org/10.1016/j.aquaculture.2017.03.005)
22. Fu J, Shen Y, Xu X, Chen Y, Li D, Li J. Multiplex microsatellite PCR sets for parentage assignment of grass carp (*Ctenopharyngodon idella*). *Aquac Int.* 2013;21:1195-1207. doi:[10.1007/s10499-013-9623-z](https://doi.org/10.1007/s10499-013-9623-z)
23. Muiocha DA, Onming S, Na-Nakorn U. Growth Performance, Genetic Diversity and Morphometric Traits of an Introduced Wild and Hatchery Population of *Clarias macrocephalus* Güther, 1864. *J Fish Environ.* 2017;41(2):1-19. <https://li01.tci-thaijo.org/index.php/JFE/article/view/92492>
24. Ye B, Wan Z, Wang L, et al. Heritability of growth traits in the Asian seabass (*Lates calcarifer*). *Aquaculture and Fisheries.* 2017;2(3):112-118. doi:[10.1016/j.aaf.2017.06.001](https://doi.org/10.1016/j.aaf.2017.06.001)

## SUPPLEMENTARY TABLES AND FIGURES

**Table S1. Basic information on *Percocypris pingi* broodstock**

| Parent ID | Total Length (cm) | Body Length (cm) | Body Weight (g) |
|-----------|-------------------|------------------|-----------------|
| ♀1        | 71.03             | 62.13            | 3074.20         |
| ♀2        | 80.08             | 72.07            | 3482.00         |
| ♀3        | 65.97             | 58.37            | 2784.50         |
| ♀4        | 61.92             | 54.62            | 2201.90         |
| ♀5        | 60.28             | 53.20            | 2053.30         |
| ♀6        | 63.86             | 56.37            | 2741.70         |
| ♂1        | 61.88             | 53.84            | 1537.00         |
| ♂2        | 53.69             | 47.58            | 1337.50         |
| ♂3        | 47.07             | 42.90            | 907.90          |
| ♂4        | 52.73             | 45.67            | 1027.70         |
| ♂5        | 48.68             | 42.86            | 817.30          |
| ♂6        | 50.28             | 44.30            | 926.70          |

**Table S2. Mating design for constructing half-sibling families using a nested design**

| Parental Line Number | ♂1    | ♂2    | ♂3    | ♂4    | ♂5    | ♂6    |
|----------------------|-------|-------|-------|-------|-------|-------|
| ♀1                   | JX 1  | JX 2  | JX 3  | JX 10 | JX 11 | JX 12 |
| ♀2                   | JX 4  | JX 5  | JX 6  | JX 13 | JX 14 | JX 15 |
| ♀3                   | JX 7  | JX 8  | JX 9  | JX 16 | JX 17 | JX 18 |
| ♀4                   | JX 19 | JX 20 | JX 21 | JX 28 | JX 29 | JX 30 |
| ♀5                   | JX 22 | JX 23 | JX 24 | JX 31 | JX 32 | JX 33 |
| ♀6                   | JX 25 | JX 26 | JX 27 | JX 34 | JX 35 | JX 36 |

Note: Horizontal axis (column): Parent number(♂1 to ♂6); Vertical axis (line): Motherbook number(♀1 to ♀6); Cells: Each cell represents a specific mating combination and is labeled with a family number (JX 1 to JX 36)

**Table S3. Fourteen half-sibling family lines of *Percocypris pingi***

| Family ID | Maternal ID | Paternal ID | Cross Type |
|-----------|-------------|-------------|------------|
| JX 4      | ♀2          | ♂1          | ♀2 × ♂1    |
| JX 5      | ♀2          | ♂2          | ♀2 × ♂2    |
| JX 6      | ♀2          | ♂3          | ♀2 × ♂3    |
| JX 13     | ♀2          | ♂4          | ♀2 × ♂4    |
| JX 14     | ♀2          | ♂5          | ♀2 × ♂5    |
| JX 15     | ♀2          | ♂6          | ♀2 × ♂6    |
| JX 28     | ♀4          | ♂4          | ♀4 × ♂4    |
| JX 29     | ♀4          | ♂5          | ♀4 × ♂5    |
| JX 30     | ♀4          | ♂6          | ♀4 × ♂6    |
| JX 31     | ♀5          | ♂4          | ♀5 × ♂4    |
| JX 32     | ♀5          | ♂5          | ♀5 × ♂5    |
| JX 33     | ♀5          | ♂6          | ♀5 × ♂6    |
| JX 34     | ♀6          | ♂4          | ♀6 × ♂4    |
| JX 36     | ♀6          | ♂6          | ♀6 × ♂6    |

**Table S4. Sample details used for full-length transcriptome and transcriptome sequencing**

| Code | Number of samples | Purpose of the sample                    | Sample source                                      | Latitude and longitude            |
|------|-------------------|------------------------------------------|----------------------------------------------------|-----------------------------------|
| YLJ  | 3                 | Full-length and transcriptome sequencing | Yalong River Stock Enhancement and Release Station | 101°38'45.258" E, 28°17'59.723" N |
| DDH  | 3                 | Transcriptome sequencing                 | Heima Fish Stock Enhancement and Release Station   | 102°80'58.29" E, 29°24'53.91" N   |
| LS   | 3                 | Transcriptome sequencing                 | Sichuan Runjie Hongda Aquatic Technology Co., Ltd. | 102°55'36.552" E, 29°56'4.664" N  |
| MS   | 3                 | Transcriptome sequencing                 | Zhougong River Ya Fish Co., Ltd.                   | 102°59'43.58" E, 30°00'14.22" N   |

**Table S5. Tissue sampling for *Percocypris pingi* in the study of RNA extraction and sequencing**

| Tissue Sample | Processing Method                   |
|---------------|-------------------------------------|
| Brain         | Combined processing for all samples |
| Gill          | Combined processing for all samples |
| Heart         | Combined processing for all samples |
| Spleen        | Combined processing for all samples |
| Kidney        | Combined processing for all samples |
| Liver         | Combined processing for all samples |
| Scale         | Combined processing for all samples |
| Skin          | Combined processing for all samples |
| Dorsal Muscle | Combined processing for all samples |

**Table S6. Sequence type statistics of full-length transcriptome of *Percocypris pingi***

| Type                    | Quantity | Min Length (bp) | Avg Length (bp) | Max Length (bp) | N50 (bp) |
|-------------------------|----------|-----------------|-----------------|-----------------|----------|
| CCS                     | 554,382  | 187             | 1584            | 13580           | 1674     |
| FLNC (Including Poly A) | 499,007  | 50              | 1412            | 7679            | 1676     |
| High-quality isoforms   | 132,504  | 50              | 1673            | 7299            | 2022     |
| Non-redundant isoforms  | 70,321   | 50              | 1882            | 7299            | 2243     |

**Table S7. Genetic diversity of nine SSR loci in the parent and offspring populations of *Percocypris pingi***

| Locus | Allele frequency |         |         |         |         |         |         |         |         |          |
|-------|------------------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
|       | Allele1          | Allele2 | Allele3 | Allele4 | Allele5 | Allele6 | Allele7 | Allele8 | Allele9 | Allele10 |
| PP021 | 0.693            | 0.028   | 0.046   | 0.005   | 0.055   | 0.123   | 0.042   | 0.008   | -       | -        |
| PP001 | 0.014            | 0.183   | 0.670   | 0.034   | 0.022   | 0.076   | 0.001   | -       | -       | -        |
| PP007 | 0.042            | 0.253   | 0.432   | 0.101   | 0.003   | 0.143   | 0.025   | 0.001   | -       | -        |
| PP028 | 0.096            | 0.089   | 0.208   | 0.239   | 0.368   | -       | -       | -       | -       | -        |
| PP034 | 0.157            | 0.093   | 0.070   | 0.263   | 0.417   | -       | -       | -       | -       | -        |
| PP016 | 0.075            | 0.302   | 0.125   | 0.446   | 0.019   | 0.033   | -       | -       | -       | -        |
| PP006 | 0.032            | 0.001   | 0.028   | 0.457   | 0.016   | 0.151   | 0.027   | 0.026   | 0.225   | 0.037    |
| PP047 | 0.072            | 0.006   | 0.08    | 0.261   | 0.31    | 0.045   | 0.004   | 0.002   | 0.127   | 0.093    |
| PP042 | 0.001            | 0.278   | 0.285   | 0.088   | 0.186   | 0.162   | -       | -       | -       | -        |

Note: Allele 1 to Allele 10 represent alleles 1 to 10 respectively



**Table S8. Basic information of twenty-two SSR Loci in *Percocypris pingi***

| Locus | Repeat motif | Annealing temperatures | Number of allele (bp) | Primer Sequence(5'~3')                                    |
|-------|--------------|------------------------|-----------------------|-----------------------------------------------------------|
| PP001 | (CA/TG)12    | 54.0                   | 279                   | F: GAATCCCTTACAACCTTACC<br>R: CGTTAAAGGAATGTCAGAAT        |
| PP005 | (CA/TG)10    | 54.0                   | 235                   | F: GCCAAAAGATGCCAAAGAAAT<br>R: TACCCTTATTGAAAACCTGACT     |
| PP006 | (AC/GT)12    | 54.0                   | 279                   | F: GAAAGCCCTAGAAAAATATTTGCT<br>R: TGGGTTTTACTGACCAGAGG    |
| PP007 | (AC/GT)10    | 54.0                   | 257                   | F: GTCTTCTGTAATCTGTAACCTGC<br>R: TGAGTAGCACAATAATCTTAAGCT |
| PP011 | (CA/TG)7     | 60.0                   | 230                   | F: GAATTTGCGCTTCACACACCGT<br>R: CTAGGTTGCAGCAGTCACATTG    |
| PP013 | (AC/GT)11    | 54.0                   | 324                   | F: GGATCAAAGTAAGAGCTGGT<br>R: GGCGGTAATTATTGCCACAT        |
| PP016 | (AC/GT)17    | 60.0                   | 327                   | F: AAGTGAAGAGCGGAGTTG<br>R: AATCCTCATCGTCTGTCTG           |
| PP021 | (AC/GT)6     | 55.0                   | 303                   | F: TATGATGCAATCAATGAATGG<br>R: CGAAACATCTGGACAGAAG        |
| PP023 | (CA/TG)10    | 55.3                   | 305                   | F: GTCCAGTTTCTGTCTCCAA<br>R: AAGCGTTTCTCCATCTCTC          |
| PP024 | (AC/GT)19    | 58.1                   | 336                   | F: CCGTGTCTCAGCAGCTGAT<br>R: CTACAGGATGCCAGGTTGG          |
| PP025 | (AG/CT)9     | 58.1                   | 338                   | F: TCTCCTCACCCAGTCTG<br>R: GGCTTCTTCACATCGTTCAC           |
| PP026 | (AC/GT)9     | 55.3                   | 327                   | F: TACCTCCCTTCTTCTCTCC<br>R: CCACAGATGCCTCTTCAAG          |
| PP028 | (CA/TG)11    | 55.3                   | 277                   | F: GTGCTACCTGTGCCATCC<br>R: GGAGGGTTAGTATGATTACACG        |
| PP030 | (CA/TG)14    | 58.1                   | 291                   | F: GAACACGCATACAGAGGG<br>R: CGCATCATCGTCGCCATC            |
| PP034 | (AC/GT)11    | 58.1                   | 280                   | F: AAGATTCCTTTCCGTCAAGAC<br>R: GTCGTCGGTGAAGTGCATC        |
| PP036 | (AC/GT)13    | 55.3                   | 344                   | F: TAAGAGACAGCAGCTTCG<br>R: CAAGTCAACTCTGTGGTAGG          |
| PP038 | (AC/GT)12    | 52.0                   | 323                   | F: AGAGTATTACAGAGAGTAACAAT<br>R: GAATAACTGGTGGTCTTAAC     |
| PP041 | (AC/GT)20    | 52.0                   | 344                   | F: AGACTGTTGACTGCTGGTT<br>R: CACTTTCAGACATTACTGAATAT      |
| PP042 | (CA/TG)11    | 55.2                   | 286                   | F: GTTTCCAGACGCTCCCTA<br>R: TTGCGCCTCAGTGTATACT           |
| PP043 | (AC/GT)15    | 55.2                   | 308                   | F: GTGGATTCTTGTCCTGG<br>R: CGTCTGTTTCTGGTTCTGG            |
| PP047 | (AC/GT)21    | 55.2                   | 284                   | F: GAAGAGATGGAGGTGGATGT<br>R: CTGATTGGTCAAGGAGGGAT        |
| PP048 | (ATC/GAT)6   | 58.4                   | 268                   | F: GACCATGATCGAGCACAGC<br>R: TCTCCATCTTGCGGCACT           |

**Table S9. Genetic diversity of nine SSR loci in the parent of *Percocypris pingi***

| Locus | Na   | Ne    | Ho    | He    | PIC   | I     | HWE   |
|-------|------|-------|-------|-------|-------|-------|-------|
| PP021 | 7    | 4.184 | 0.917 | 0.772 | 0.731 | 1.641 | ns    |
| PP001 | 6    | 2.549 | 1.000 | 0.646 | 0.558 | 1.067 | **    |
| PP007 | 8    | 4.521 | 0.917 | 0.761 | 0.749 | 1.714 | **    |
| PP028 | 5    | 4.235 | 1.000 | 0.776 | 0.728 | 1.525 | *     |
| PP034 | 5    | 3.588 | 1.000 | 0.729 | 0.671 | 1.379 | ns    |
| PP016 | 6    | 4.451 | 1.000 | 0.802 | 0.744 | 1.622 | **    |
| PP006 | 9    | 4.247 | 1.000 | 0.802 | 0.740 | 1.754 | ns    |
| PP047 | 6    | 3.065 | 1.000 | 0.655 | 0.626 | 1.335 | ns    |
| PP042 | 5    | 4.345 | 1.000 | 0.788 | 0.732 | 1.527 | *     |
| AVE   | 6.33 | 3.909 | 0.981 | 0.748 | 0.697 | 1.507 | 0.137 |

Note: "ns" indicates compliance with Hardy-Weinberg equilibrium, and "\*" indicates a significant deviation from Hardy-Weinberg equilibrium after Bonferroni correction ( $P < 0.05$ ), "\*\*\*" indicates a highly significant deviation ( $P < 0.01$ )

**Table S10. Exclusion probabilities and cumulative exclusion probabilities for parentage identification in *Percocypris pingi***

| Locus | E-1P  | E-2P  | E-PP  | CE-1P    | CE-2P    | CE-PP    |
|-------|-------|-------|-------|----------|----------|----------|
| PP021 | 0.378 | 0.560 | 0.753 | 0.378014 | 0.559749 | 0.752550 |
| PP001 | 0.167 | 0.316 | 0.480 | 0.481597 | 0.698877 | 0.871307 |
| PP007 | 0.403 | 0.582 | 0.771 | 0.690333 | 0.874134 | 0.970490 |
| PP028 | 0.366 | 0.546 | 0.730 | 0.803671 | 0.942880 | 0.992033 |
| PP034 | 0.301 | 0.472 | 0.648 | 0.862702 | 0.969814 | 0.997192 |
| PP016 | 0.391 | 0.571 | 0.758 | 0.916357 | 0.987055 | 0.999320 |
| PP006 | 0.374 | 0.559 | 0.761 | 0.947639 | 0.994289 | 0.999837 |
| PP047 | 0.272 | 0.445 | 0.635 | 0.961890 | 0.996832 | 0.999941 |
| PP042 | 0.370 | 0.549 | 0.729 | 0.975994 | 0.998571 | 0.999984 |

Note:

E-1P: Non-exclusion probability at a single locus when one parent is unknown.

CE-1P: Combined exclusion probability across multiple loci when one parent is unknown.

E-2P: Non-exclusion probability at a single locus when one parent is known.

CE-2P: Combined exclusion probability across multiple loci when one parent is known.

E-PP: Non-exclusion probability at a single locus when both parents are unknown.

CE-PP: Combined exclusion probability across multiple loci when both parents are unknown.

**Table S11. Bruvo's genetic distances of the 12 parents of *Percocypris pingi***

| Family ID | Parent Pair | Bruvo's Genetic Distance |
|-----------|-------------|--------------------------|
| JX4       | ♀2×♂1       | 0.373155382              |
| JX5       | ♀2×♂2       | 0.377466838              |
| JX6       | ♀2×♂3       | 0.360527886              |
| JX13      | ♀2×♂4       | 0.334120009              |
| JX14      | ♀2×♂5       | 0.396692912              |
| JX15      | ♀2×♂6       | 0.333496094              |
| JX28      | ♀4×♂4       | 0.380208333              |
| JX29      | ♀4×♂5       | 0.366746691              |
| JX30      | ♀4×♂6       | 0.38156467               |
| JX31      | ♀5×♂4       | 0.388726128              |
| JX32      | ♀5×♂5       | 0.394958072              |
| JX33      | ♀5×♂6       | 0.398191664              |
| JX34      | ♀6×♂4       | 0.362792969              |
| JX36      | ♀6×♂6       | 0.361950345              |

**Table S12. Descriptive statistics of three growth traits in a mixed breeding family of *Percocypris pingi***

| Trait           | Mean  | SD   | Maximum | Minimum | CV/%  |
|-----------------|-------|------|---------|---------|-------|
| Total length/cm | 12.83 | 1.25 | 15.73   | 9.70    | 9.71  |
| Body length/cm  | 10.82 | 1.14 | 13.46   | 7.71    | 10.51 |
| Body weight/g   | 17.48 | 4.27 | 27.23   | 8.28    | 24.41 |

**Table S13. Genetic and phenotypic correlations between traits**

| Trait           | Total Length/cm  | Body Length/cm   | Body Weight/g    |
|-----------------|------------------|------------------|------------------|
| Total Length/cm |                  | 0.976(0.003) *** | 0.935(0.009) *** |
| Body Length/cm  | 0.988(0.016) *** |                  | 0.933(0.011) *** |
| Body Weight/g   | 0.965(0.050) *** | 0.921(0.080) *** |                  |

Note: Above the diagonal: Genetic correlation.; Below the diagonal: Phenotypic correlation. In parentheses: Standard Error (SE); P-value: obtained through LRT test, such as "\*" indicating  $P < 0.05$ , "\*\*\*" indicating  $P < 0.01$ , "\*\*\*\*" indicating  $P < 0.001$ , and " " indicating  $P > 0.05$

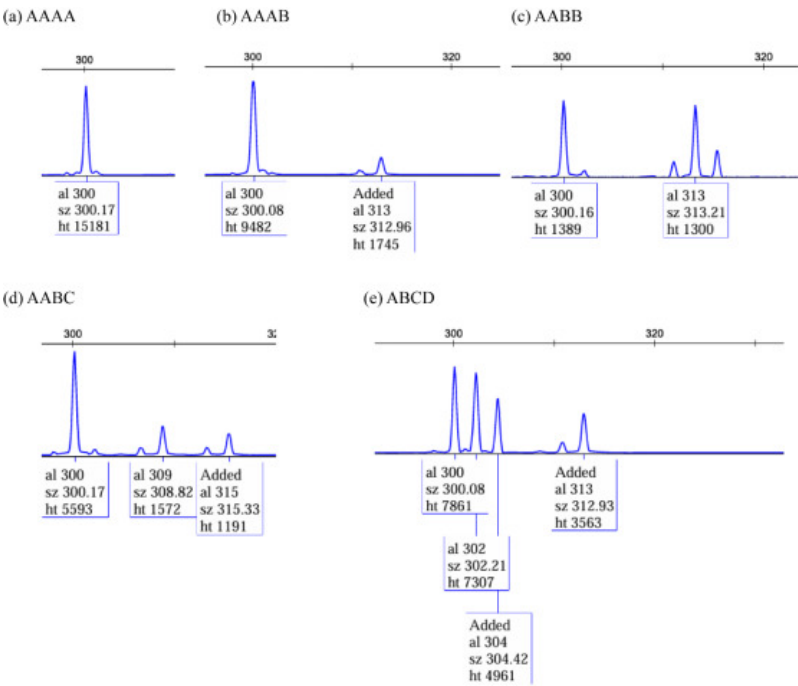


Fig. S1. A schematic representation of allele discrimination in *Percocypris pingi* based on microsatellite genotyping. Panels illustrate five typical allelic combinations: (a) AAAA, (b) AAAB, (c) AABB, (d) AABC, and (e) ABCD, representing increasing levels of allelic variation detected across loci.

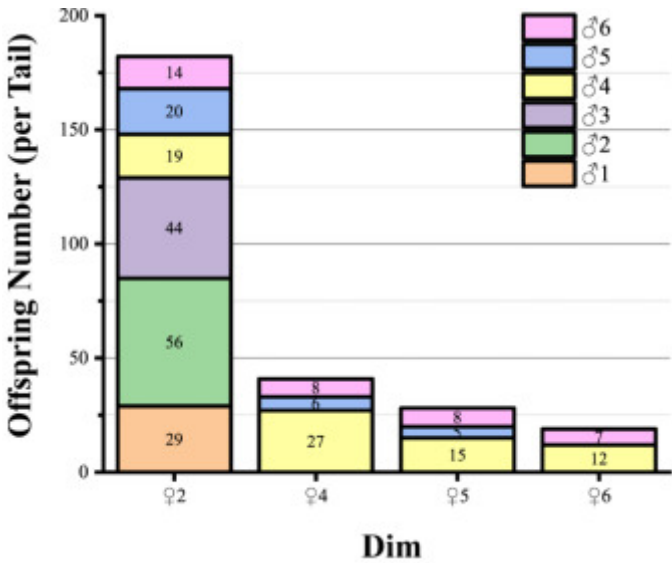


Fig. S2. Distribution of offspring numbers corresponding to different maternal and paternal genotypes in the mixed-rearing group of *Percocypris pingi*. The figure illustrates the number of offspring assigned to each parental genotype combination under communal rearing conditions.