BMC Genomics

RESEARCH

Open Access

Identification of sex-specific markers using genome re-sequencing in the blunt snout bream (*Megalobrama amblycephala*)

Yuye Fu[1], Lifei Luo[1*], Shilong Wang[1], Yue Yu[1], Yao Wang[1] and Zexia Gao[1,2,3*]

Abstract

Background The blunt snout bream (*Megalobrama amblycephala*) is an important economic freshwater fish in China with tender flesh and high nutritional value. With the cultivation of superior new varieties and the expansion of breeding scale, it becomes imperative to employ sex-control technology to cultivate monosexual populations of *M. amblycephala*, thereby preventing the deterioration of desirable traits. The development of specific markers capable of accurately identifying the sex of *M. amblycephala* would facilitate the determination of the genetic sex of the breeding population before gonad maturation, thereby expediting the processes of sex-controlled breeding of *M. amblycephala*.

Results A whole-genome re-sequencing was performed for 116 females and 141 males *M. amblycephala* collected from nine populations. Seven candidate male-specific sequences were identified through comparative analysis of male and female genomes, which were further compared with the sequencing data of 257 individuals, and finally three male-specific sequences were generated. These three sequences were further validated by PCR amplification in 32 males and 32 females to confirm their potential as male-specific molecular markers for *M. amblycephala*. One of these markers showed potential applicability in *M. pellegrini* as well, enabling males to be identified using this specific molecular marker.

Conclusions The study provides a high-efficiency and cost-effective approach for the genetic sex identification in two species of *Megalobrama*. The developed markers in this study have great potential in facilitating sex-controlled breeding of *M. amblycephala* and *M. pellegrini*, while also contributing valuable insights into the underlying mechanisms of fish sex determination.

Keywords *Megalobrama amblycephala*, Whole-genome re-sequencing, Genetic sex identification, Molecular marker

*Correspondence:
Lifei Luo
luolifei@mail.hzau.edu.cn
Zexia Gao
gaozx@mail.hzau.edu.cn
[1]College of Fisheries, Hubei Hongshan Laboratory / Key Lab of Freshwater Animal Breeding, Ministry of Agriculture and Rural Affairs / Engineering Research Center of Green development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
[2]Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China
[3]Engineering Technology Research Center for Fish Breeding and Culture in Hubei Province, Wuhan 430070, China

Background

Fish is a diverse group, playing a pivotal role in the evolutionary history of vertebrates. Currently, over 32,000 species are known, accounting for more than half of all vertebrate species [1]. Fish sex determination is complex and diverse, covering almost all known sex determination types of vertebrates, and there are unique sex differentiation methods such as natural sex reversal [2, 3]. Therefore, the study of fish sex is of great significance for elucidating the sex-determination mechanism of vertebrates. At the same time, numerous fish species exhibit significant sexual dimorphism in important economic traits such as growth and body size [4, 5], sex-control in fish has emerged as a pivotal domain of genetic breeding research, playing a crucial role in the breeding of superior fish species and the maintenance of superior traits. Furthermore, the gonads or female gametes of certain fish, such as sturgeon eggs, hold substantial economic value. Therefore, the cultivation of faster-growing or mono-sex aquaculture strains with specific gonadal characteristics holds significant economic importance. For some species facing the decline of germplasm resources, as well as some fish, such as Nile tilapia (*Oreochromis niloticus*) and Atlantic salmon (*Salmo salar*), before reaching the market size, they often appear to mature in advance and enter the breeding period, which in turn affects growth. Cultured unisexual populations can effectively avoid reproductive activities and protect germplasm to a certain extent [6]. Some fish species, such as orange spotted grouper (*Epinephelus coioides*) and Asian swamp eel (*Monopterus albus*), are hermaphroditic and capable of natural sex transformation, mono-sex aquaculture is essential for the management of their breeding broodstock. However, the degree of sexual chromosome differentiation is generally low in most fish species, making it difficult to identify based solely on chromosomal morphology [7]. The development of sex-specific molecular markers and the advancements in sex-control biotechnology provide an important technical approach to increase fish yield and its economic value, and lay a foundation for deciphering the molecular mechanism of fish sex determination.

Sex-specific markers are effective molecular tools for detecting genetic sex and illuminating the genetic structure of distinct sexes within species [8]. To date, a variety of techniques have been used to develop sex-specific markers, such as randomly amplified polymorphic DNA (RAPD) [9–11], amplified fragment length polymorphism (AFLP) [12, 13], simple sequence repeats (SSR) [14–16], and single nucleotide polymorphisms (SNP) [17–19]. In recent years, the development strategies of molecular markers based on high-throughput sequencing have gained popularity as a result of the ongoing maturation of sequencing technology and the declining cost of sequencing [20]. Particularly with the rise in the number of species with known genome sequences, whole-genome re-sequencing is emerging as one of the key resources for investigating the molecular breeding of plants and animals [21, 22]. The utilization of high-throughput sequencing technology has proven to be highly effective in comparing the genomic differences between male and female individuals, making it a powerful method for identifying molecular markers of sex in fish. Currently, this method has successfully identified molecular markers of sex in various fish species, including snakehead (*Channa argus*) [23], grass carp (*Ctenopharyngodon idella*) [24], large yellow croaker (*Larimichthys crocea*) [19], bighead carp (*Hypophthalmichthys molitrix*) [25], redtail catfish (*Mystus wyckioides*) [26], Pacific bluefin tuna (*Thunnus orientalis*) [27], Taiwan snakehead (*Channa maculata*) [28], and mandarin fish (*Siniperca chuatsi*) [29]. The identification of these molecular markers provides a fundamental basis for determining the genetic sex of fish and facilitating the breeding of superior new varieties through sex-control breeding techniques. Furthermore, the utilization of whole-genome re-sequencing technology has significantly improved the efficiency of sex-specific molecular markers development.

Blunt snout bream (*Megalobrama amblycephala*), naturally distributed in the middle and lower reaches of the Yangtze River and commonly known as Wuchang bream in China, is a typical herbivorous Cyprinidae fish that exhibits exceptional traits such as delicious meat, fast growth and low breeding cost [30]. Its annual production reached approximately 738,000 tons by 2023, ranking the seventh among the major cultivated freshwater fish in China [31]. Previously, researchers have developed excellent varieties of fast-growing, disease-resistant *M. amblycephala* such as 'Pujiang No. 1', 'Huahai No. 1', and 'Pujiang No. 2' [32]. However, the germplasm of *M. amblycephpara* may decline due to precocious puberty and differentiation of superior traits during long-term culture and unplanned breeding [33]. One ideal solution to these problems is to generate monosexual or sterile populations with dominant traits through sex-control technology [34, 35]. Nevertheless, most fish male and female lack heterotypic sex chromosomes, and their small morphological differences are insufficient for accurate sex discrimination [36, 37].

M. amblycephala has a male heterogametic (XX/XY) sex-determining system [38]. Although Wen et al. have developed a male-specific molecular markers of *M. amblycephala* [39], its sex-determining gene is still unclear. Therefore, the development of additional sex markers is imperative in order to enhance the discovery of genome-wide sex determining genes. If we can develop specific markers that can accurately identify the sex of *M. amblycephala*, we will be able to identify the genetic

sex of the cultured population before its gonads mature, carry out unisexual culture, effectively eliminate excessive hybridization within the population, and promote the protection of its germplasm resources. Utilizing genetic sex identification can assist farmers in promptly identifying the genetic sex of fish at an early stage, maintaining control over the sex ratio of the broodstock group at any time, and adjusting the number of parent fish in cultivation to reduce breeding costs.

In this study, three new male-specific markers were identified by re-sequencing of *M. amblycephala* population and a comparative analysis of male and female genomes. The identified sex-specific makers were also tested in other genus *Megalobrama* species. The findings will provide a theoretical basis for understanding the sex determination system of *M. amblycephala*, and contribute to its sex-controlled breeding.

Methods

Experimental fish, sample collection and DNA extraction

The experimental fish were sacrificed by anesthetized with MS-222 at 100 mg/L before fin collection and dissection. Then we performed a dissection of the gonads while the fish were still anesthetized to determine their sex. In the processes of sex-specific molecular marker screening, a total of 257 *M. amblycephala* (116 females and 141 males), nine populations from Yuni Lake (YNL, $n=30$), Poyang Lake (PYL, $n=29$), Liangzi Lake (LZL, $n=24$), Hubei Zhenglong Aquatic Seed Industry Co., Ltd. (ZL, $n=29$), 'Huahai No. 1' (HH1, $n=29$), 'Pujiang No. 1' (PJ1, $n=29$), 'Pujiang No. 2' (PJ2, $n=28$), Guangshui (GS, $n=29$), and Qianjiang (QJ, $n=30$), were collected. Additionally, a total of 32 males and 32 females *M. amblycephala* were collected from Ezhou (EZ) breeding base for the verification of sex-specific markers. The species in *Megalobrama* genus, including *M. terminalis*, *M. hoffmanni*, and *M. pellegrini* were collected from Hangzhou, Guangzhou, and Yibin, respectively. The fin tissues from all samples were preserved in 95% ethanol and stored at -80°C. The sex of *M. amblycephala* was primarily determined through anatomical observation of the gonads, and individuals that could not be fully tested were identified by gonadal sections. Specifically, the gonads were treated with Bouin's solution, then embedded in paraffin for sectioning, and finally stained with hematoxylin and eosin (HE). Genomic DNA was isolated using ammonium acetate extraction and determined by NanoDrop2000 (Thermo Scientific, USA) and 1% agarose gel electrophoresis. The DNA was stored at -80 °C for subsequent experiments.

Whole-genome re-sequencing

DNA samples from YNL, PYL, LZL, ZL, HH1, PJ1, PJ2, GS, and QJ populations were used for libraries construction. Genome re-sequencing DNA libraries were constructed using the YZSeq™ Tn5 Library Prep Kit (YINGZI GENE, Wuhan, China) following the manufacturer's instructions. Briefly, the libraries were first prepared using Tn5 transposase, and with the processes of DNA fragmentation, end-polishing, and adaptors ligation. The linker self-ligated fragments were removed by screening with magnetic beads. The libraries templates were enriched by PCR amplification, and finally the PCR product was recovered by magnetic beads to obtain the final libraries. Finally, a total of 116 female and 141 male individual libraries were sequenced on the MGISEQ-T7 platform with PE150 mode. The raw reads obtained by sequencing were quality-controlled using Fast QC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Clean reads were obtained after removing the adaptor sequences. Trimmomatic v0.36 [40] was used to remove paired reads containing more than 10% ambiguous bases or more than 50% low-quality bases (Phred Score < 5).

Screening of sex-specific markers

To screen candidate Y-specific segments of *M. amblycephala*, male-specific sequences were identified using differential subtraction comparative analysis of male and female genomes. Firstly, the clean data of 116 females and 141 males were aligned to the female reference genome (GCF_018812025.1) of *M. amblycephala* using BWA (v0.7.12) [41], and any sequences that did not match the female reference genome were extracted. These extracted sequences may include: shared sequences between females and males that do not match the reference genome; sequence variations among individual females; sequence variations among individual males; and male-specific sequences. Secondly, the sequences that could not be aligned with the female reference genome were extracted from the above 10 high-depth male samples, and a fragment set containing male-specific fragments was assembled using SPAdes (v3.13.0) [42]. Thirdly, the re-sequencing reads of 20 female individuals were randomly selected and bidirectional comparison was conducted with the assembly results obtained in the second step, and the corresponding sequences were removed. Fourthly, re-sequencing reads from another set of 20 male individuals were randomly selected for bidirectional comparison with filtered results from the third step, to retain sequences that could be aligned by all male individuals, thereby identifying candidate male-specific sequences. Finally, the shorter fragments (<150bp) were removed, and the remaining fragments were identified as male-specific fragments for further verification.

Validation of candidate sex-specific markers

Primer pairs were designed for each of the screened sex-specific markers. Subsequently, the primer pairs were validated by PCR in EZ breeding population, Hubei Province. The *β-actin* gene was used as a positive control (Forward primer: ACCCACACTGTGCCCATCTA; Reverse primer: CGGACAATTTCCCTCTCAGCTG). In addition, sex-specific marker and *β-actin* primers were added to the same reaction system for PCR amplification. The PCR was performed in a total volume of 10 μL: 5 μL 2× Hieff PCR Master Mix (Yeasen, Shanghai, China), 0.25 μL forward and reverse primers mix of each pair (10 mol/L each), 0.5 μL template DNA (100 ng/μL), and 3.5 μL ddH_2O. The PCR amplification conditions were as follows: 94℃ for 5 min; 30 cycles of 94℃ for 30 s, 56℃ for 30 s and 72℃ for 30 s; then 72℃ for 5 min. The PCR products were analyzed by 1.5% agarose gel electrophoresis and sequenced by Wuhan Tianyi Huayu Gene Technology Co., Ltd. (Wuhan, China).

Annotation of sex-specific molecular marker fragments

The sequencing depth of the re-sequencing ranging from 10× to 23×, is insufficient for the male reference genome assembly. Consequently, accurate localization of above male-specific marker fragments becomes unattainable. To determine whether these molecular markers are situated in gene regions, we performed annotation based on comparison with the NCBI database (https://www.ncbi.nlm.nih.gov/).

Phylogenetic tree construction

All protein sequences were collected from genome annotation of 10 species of Cyprinidae fish, including *M. amblycephala* (Genome ID: GCF_018812025.1), *M. hoffmanni*, *M. terminalis*, *M. pellegrini*, *Parabramis pekinensis* (The genome of the four species were sequenced and assembled by our lab, unpublished data), *Elopichthys bambusa* (Genome ID: GCA_037101425.1), *Ancherythroculter nigrocauda* (Genome ID: GCA_036281575.1), *Hypophthalmichthys nobilis* (https://ftp.cngb.org/pub/CNSA/data3/CNP0000974/CNS0209134/CNA0019189/HN.genomic.fasta.gz), *Hypophthalmichthys molitrix* (Genome ID: GCA_037950675.1), and *C. idella* (Genome ID: GCA_019924925.1). OrthoFinder (v2.5.4) (https://github.com/davidemms/OrthoFinder) was used to cluster gene families of 10 species and obtain single-copy orthologous gene families. Only genes with an amino acid length of 100 or greater were retained by filtering the shared single-copy orthologous gene families. MUSCLE (v3.8.31) [43] was used to perform multiple sequence alignment for genes within each single-copy orthologous gene family. Subsequently, the multiple sequence alignment results were combined and converted into super-gene alignment in PHYLIP format. Finally, RAxML (v8.2.12) [44] was used to construct the evolutionary tree based on the maximum likelihood method.

Results

Phenotypic sex identification

The sex of most individuals was identified through anatomical observation of the gonads (Fig. 1A-B). Furthermore, the histological section showed that ovarian follicles were clearly visible in the ovaries of those female individuals, while spermatozoa were clearly visible in the testis of those male individuals (Fig. 1E-F). For individuals whose sex cannot be identified by the appearance of the gonads, we employed paraffin-embedded gonads and utilized HE staining to observe the cell structure to identify the sex (Fig. 1C-D). The results of gonadal tissue

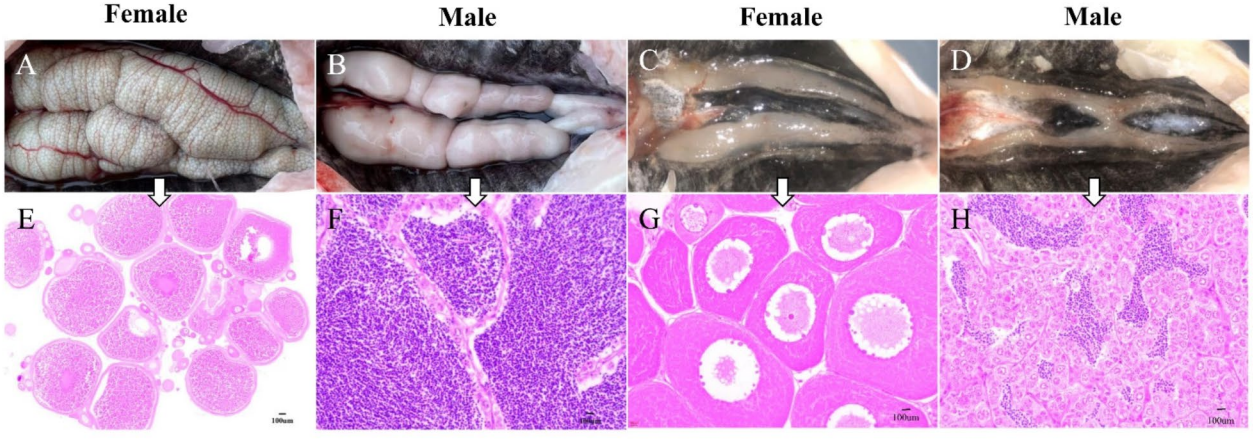

Fig. 1 Identification of phenotypic sex in blunt snout bream. **(A)-(D)** Gonad morphology of female and male individuals. **(A)-(B)** Individuals whose sex can be identified through anatomical observation; **(C)-(D)** Individuals whose sex cannot be identified through anatomical observation. **(E)-(H)** Sex identification of blunt snout bream by staining gonad tissue slices with hematoxylin and eosin. **E**, female, visible mature follicles can be observed in the ovary; F, male, spermatozoa were clearly visible in the testis; **G**, oocytes; **H**, spermatocytes

sections indicated that individuals with oocytes observed in the sections were females, while individuals with many spermatocytes observed in the sections were males (Fig. 1G-H). This enables precise determination of their phenotypic sex. After conducting anatomical observations and examining gonad tissue slices, it was confirmed that there were 116 females and 141 males among the 257 individuals used for genome re-sequencing.

Re-sequencing data analysis and candidate sex-specific tags screening

The raw reads obtained by re-sequencing of female and male individuals in nine populations ranged from 67,271,000 to 272,520,042, and the clean reads ranged from 67,270,910 to 272,519,954 were generated subsequently (Table S1). In addition, Q20 and Q30 of the re-sequencing data ranged from 90.12 to 98.52% and 80.22–95.07%, respectively. After aligning the clean data of 116 females and 141 males to the female reference genome of *M. amblycephala* (GCF_018812025.1), the sequences aligned with reference genomes and differential sequences between female individuals were removed. Then, 10 high-depth male samples were assembled based on the remaining sequences, and 262,853 sequences were obtained, including male-specific sequences and sequences that exist in both males and females but not aligned to the female reference genome. Subsequently, 163,710 sequences were obtained after removing both male and female reference genome sequences. Finally, 45 candidate male-specific tag sequences were obtained by filtering the male individual difference sequences, of which seven tag sequences greater than 150 bp were used as the final candidate male-specific tag sequences for subsequent verification (Fig. 2).

Validation and annotation of sex-specific markers

The seven final candidate male-specific tag sequences were compared with the sequencing data of 257 *M. amblycephala* individuals and further screened to obtain three candidate fragments, which were identified as candidate sex-specific markers (Table 1). The amplification results of the optimal primers for each fragment in 12 female and 12 male *M. amblycephala* showed that the optimal primers of M-marker 1, M-marker 2 and M-marker 3 amplified bands of 173 bp, 275 bp, and 436 bp in male individuals, respectively (Table 2; Fig. 3), while no bands were amplified in female individuals. Following sequencing, it was discovered that the bands amplified in males did not match the female reference genome and shared the same nucleotide sequences as the reference amplified sequences of the corresponding primer pairs. To further verify the accuracy of the male-specific markers, 32 males and 32 females were randomly selected from the EZ breeding population for PCR. The results demonstrated that three molecular marker primers could only amplify bands in male individuals when the positive controls verified that all male and female individual DNA could be effectively used for PCR amplification (Fig. 4 and Fig. S1-S3), which strongly supports the validity of the three sex-specific markers for accurately identifying the genetic sex of *M. amblycephala*. Annotation of three male-specific molecular marker fragments showed that partial sequence of the M-marker 1 could be aligned with the LOC127516645, *trim9* and *cacng6a* genes of *C. idella*, and partial sequence could be aligned with the *trim9* gene of *M. amblycephala* (Table S2). The partial sequence of M-marker 2 was aligned with LOC125259710 gene of *M. amblycephala*, *bglap* and LOC127499833 genes of *C. idella*, and the partial sequence of M-marker 3 was aligned to the non-coding RNA LOC131524810 of *Onychostoma macrolepis* and the *lmx1al* gene of *C. idella* (Table S2).

The application of sex-specific markers in other closely related species

M. amblycephala, *M. terminalis*, *M. hoffmanni*, and *M. pellegrini* all belong to the *Megalobrama* genus. The maximum likelihood method was employed to construct phylogenetic tree from 2,915 single-copy orthologous gene families (Table S3 and S4), revealing that the four species were closely related (Fig. 5A). M-marker 1 and M-marker 2 could amplify 173 bp and 275 bp bands in some male and female individuals of *M. terminalis* and *M. hoffmanni*, but not in both male and female individuals of *M. pellegrini*, which could not effectively distinguish the male and female individuals of the three species (Fig. 5B and Fig. S4). M-marker 3 amplified a 436 bp band only in male individuals of *M. pellegrini*, but not in female individuals, and this was strongly verified by the expanded population experiment (Fig. 5B-C and Fig. S4-S5).

Discussion

The discovery of sex-specific markers is crucial for elucidating sex determination in fish. It aids in identifying sex-determining regions, facilitates the identification of sex-determining genes, and enhances our understanding of the origin and evolution of sex chromosomes [45]. While most fish lack visually heteromorphic sex chromosomes, some have developed sex-determining regions or possess sex-determining genes. Identifying sex-specific markers serves as the initial step in discovering these regions and genes, thereby aiding in the inference of their sex-determining systems [24, 46, 47]. Before the mature gonadal development of *M. amblycephala*, distinguishing the outward morphology of males and females is challenging. Hence, the development of sex-specific markers for *M. amblycephala* can assist in determining the genetic sex of breeding populations before gonadal

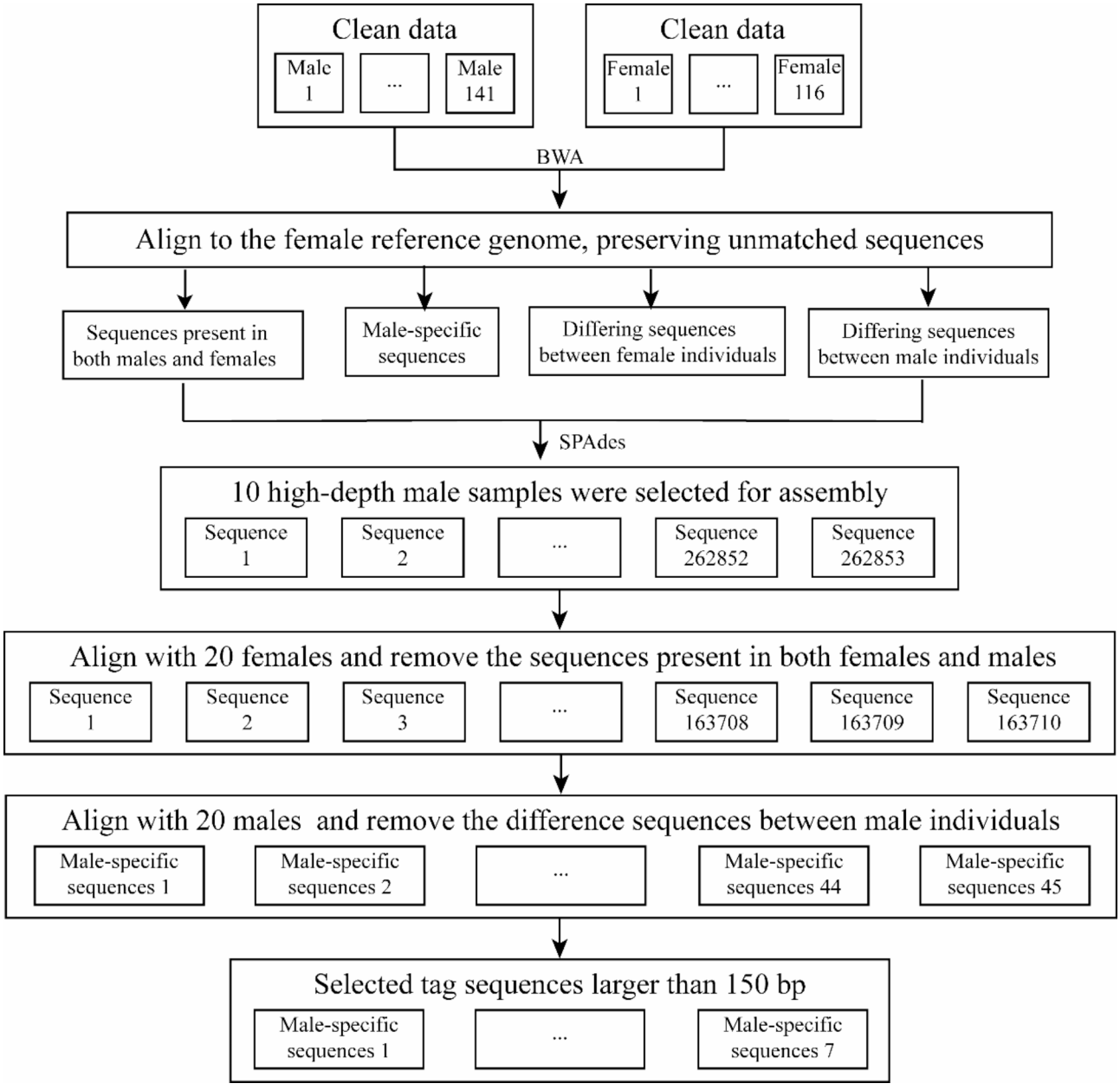

Fig. 2 Screening processes for male-specific fragments

Table 1 Mean sequence depth statistics for seven contigs in male and female populations

Contig ID	Mean sequence depth in males	Mean sequence depth in females
NODE_15451_length_647_cov_37.393312	2.1563×	0.0177×
NODE_178500_length_195_cov_72.426136	7.2787×	0.0139×
NODE_2507_length_1650_cov_40.518043	3.0348×	0.0037×
NODE_59997_length_321_cov_69.221854	4.6291×	0.0146×
NODE_26104_length_488_cov_24.123667	1.8895×	0.0027×
NODE_49161_length_355_cov_45.377976	2.8404×	0.0019×
NODE_8326_length_900_cov_44.861521	4.1431×	0.0253×

Table 2 Information of three male-specific candidate contigs and their sex-specific primers

Contig ID	Length (bp)	Marker name	Primer sequences (5′-3′)	Tm (℃)	Product length (bp)
NODE_178500_length_195_cov_72.426136	195	M-marker 1	F: GACGTCCATATGTTCAGCCA	56	173
			R: CAAGTCAGAATTGATTACAATAAAT		
NODE_59997_length_321_cov_69.221854	321	M-marker 2	F: ATCAGTAAATATGCAGAGCGAGA	56	275
			R: AGTTTCTGTCAAGTCCTCGCA		
NODE_26104_length_488_cov_24.123667	488	M-marker 3	F: AAGCAAATGCTGAAATGTTGTATA	56	436
			R: TAAAAAGTTTAAGTTTTGGCCTG		

Fig. 3 Sequences corresponding to the three sex-specific markers. The positions marked by black arrows are the forward and reverse primers. **(A)** Sequences amplified by M-marker 1; **(B)** Sequences amplified by M-marker 2; **(C)** Sequences amplified by M-marker 3

maturity for unisexual breeding. It also helps farmers control the sex ratio of the reserved parental population at any time and further regulates the number of parents to be bred. In this study, we successfully developed three PCR-based sex-specific molecular markers for *M. amblycephala* through genome re-sequencing, and verified the specificity and universality of these markers in breeding populations, providing an important molecular tool for the study of sex determination in fish.

This study presents a rapid and efficient method for developing sex-specific molecular markers using next-generation sequencing combined with bioinformatics analysis. In recent years, although the restriction site associated DNA sequencing (RAD-seq) method has been successfully applied to the development of sex markers

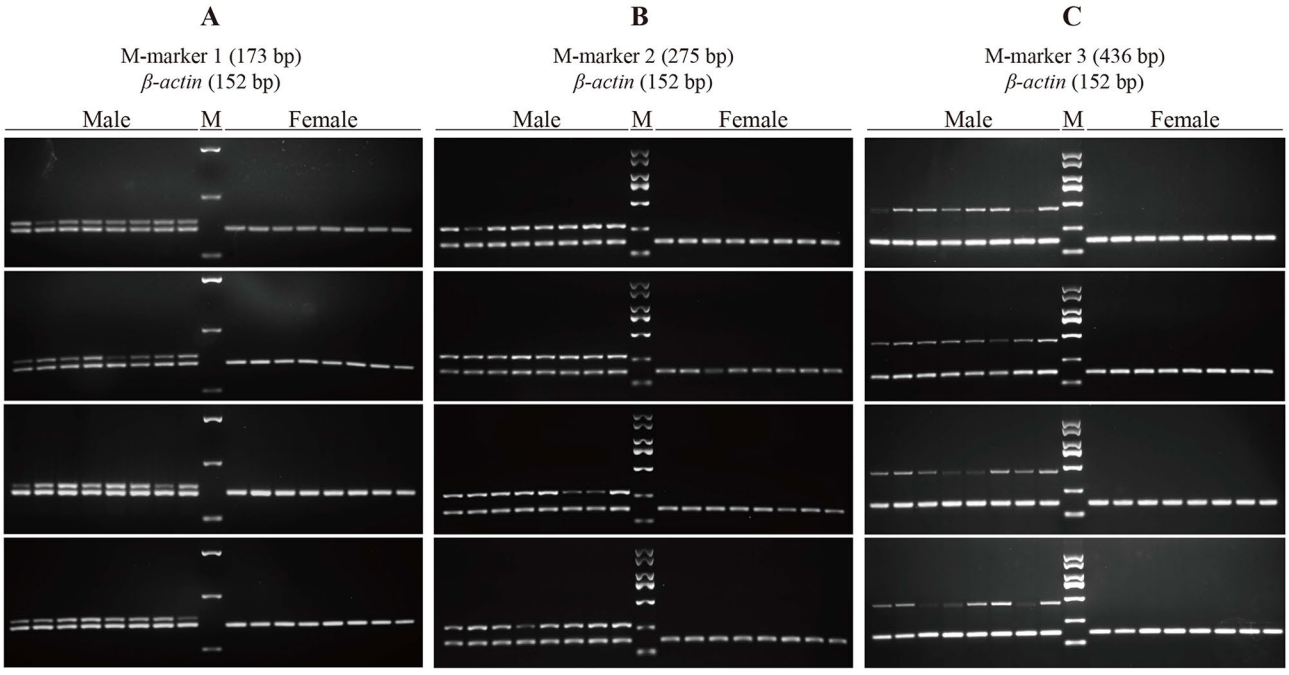

Fig. 4 PCR amplification results of three sex-specific primer pairs in EZ breeding population of blunt snout bream. Sex-specific primers and *β-actin* primers were incorporated into the same PCR reaction system, and the validity of the amplification results was verified by the presence or absence of control bands (152 bp). **(A)** Amplification of M-marker 1 resulted in a male-specific band at 173 bp. **(B)** Amplification of M-marker 2 resulted in a male-specific band at 275 bp. **(C)** Amplification of M-marker 3 resulted in a male-specific band at 436 bp. M, DL 2000 DNA marker

in species such as Lanzhou catfish (*Silurus lanzhouensis*) [48], goldfish (*Carassius auratus*) [49], Ussuri catfish (*Pseudobagrus ussuriensis*) [50], and Chinese longsnout catfish (*Leiocassis longirostris*) [51], its sequencing fragments are short and some restriction sites are often missed. And subsequent analysis can be easily affected by repeated sequences, leading to increased difficulty and cost in identifying sex-specific markers [37, 52]. In contrast, this study utilized genome re-sequencing to obtain high-throughput genomic data from multiple female and male individuals, and employed bioinformatics analysis to identify differential genomic regions between the sexes. Compared to RAD-seq, this method offers extensive coverage and can efficiently gather more comprehensive variation information, allowing for the accurate identification of sex-specific molecular markers in fish [25]. The sex-specific markers obtained by this method can be used to identify the genetic sex through simple PCR amplification and electrophoresis, which provides a simple, efficient, and reliable genetic sex identification technique for the breeding of *M. amblycephala*. Additionally, sex-controlled breeding technology combining sex molecular marker identification and sex hormone-induced reversal can quickly and effectively cultivate mono-sex population [53, 54], which could contribute to safeguarding the industrial autonomy of new germplasms or varieties of *M. amblycephala* such as those without intermuscular bones [55].

Wen et al. carried out the whole-genome re-sequencing of female and male individuals in two *M. amblycephala* populations, and assembled the male reference genome [39]. They aligned re-sequenced reads of females and males to the male reference genome and a male-specific molecular marker of *M. amblycephala* was obtained. The difference of our method is that, firstly, we carried out re-sequencing of nine *M. amblycephala* populations, covering all the new varieties and wild populations of *M. amblycephala* in China, and the molecular markers developed based on these populations are more universal. Secondly, the number of female and male individuals used to develop sex markers is larger, which can effectively remove the influence of individual pan-genomic sequence differences. Finally, we did not rely on the male reference genome in the processes of screening sex markers. We only extracted the sequences of 10 high-depth male samples that did not match the female reference genome, and assembled them into potential male specific fragments. Then, female and male individuals were used to map to the potential male specific fragments to remove the sequence variations among individuals. Therefore, our method can effectively remove the effects of individual difference sequences and quickly screen out the real sex-specific fragments.

Moreover, *M. amblycephala*, *M. terminalis*, *M. hoffmanni*, and *M. pellegrini* all belong to the *Megalobrama* genus. The phylogenetic tree of several *Megalobrama*

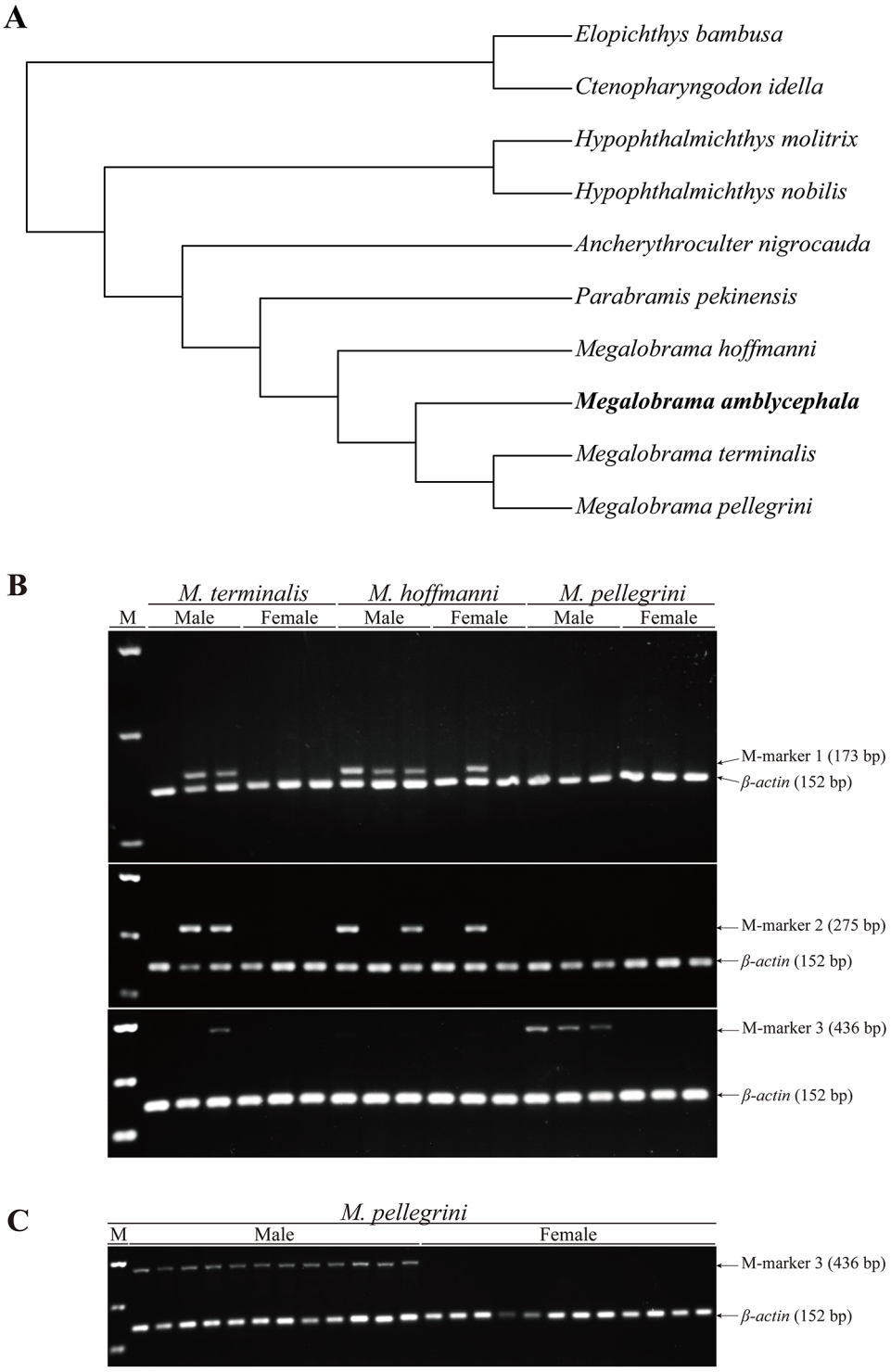

Fig. 5 Phylogenetic tree construction and PCR amplification results of three sex-specific primer pairs in other closely related species (*M. terminalis*, *M. hoffmanni*, and *M. pellegrini*). **(A)** A phylogenetic tree generated using single-copy orthologous genes. **(B)** Correspond to the results amplified by primer pairs of M-marker 1, M-marker 2 and M-marker 3. **(C)** Extended population verification of the sex-specific primer pair of M-marker 3 in *M. pellegrini* collected from Yibin. The amplified fragment of *β-actin* was 152 bp for the positive control, confirming the effectiveness of the amplification. M, DL 2000 DNA marker

species showed that the four species were closely related, and the relationship between *M. amblycephala* and *M. hoffmanni* was the closest. Therefore, it was speculated that they may have similar sex-determination regions. However, it appears that only one out of the three sex markers (M-marker 3) was conserved and only in *M. pellegrini*. This may be because the marker sequence has the highest similarity between *M. amblycephala* and *M. pellegrini*.

Three male-specific molecular markers, M-marker 1, M-marker 2, and M-marker 3, were not annotated to the well-known master sex determining genes such as *Dmrt1*, *Amh*, *Sox3*, *Gdf6*, and *Amhr2* [56]. They were annotated to the LOC127516645 gene of *C. idella*, the LOC125259710 gene of *M. amblycephala*, and the non-coding RNA LOC131524810 of *O. macrolepis*, respectively. The functions of these genes and non-coding RNAs are unknown, and whether these genes and non-coding RNAs play an important regulatory role in the sex determination of *M. amblycephala* remains to be further studied. It is often observed that sex-specific markers are developed, but sex-determining genes cannot be identified [57]. For instance, the blotted snakehead (*Channa maculata*), was found to have 23 male-specific sequences, of which six male-specific markers were effectively confirmed. Nevertheless, no potential sex-determining genes were identified [58]. A similar situation was found in Atlantic cod (*Gadus morhua*) [59], clearhead icefish (*Protosalanx hyalocranius*) [60], and largemouth bass (*Micropterus salmoides*) [61]. Undoubtedly, sex-specific markers continue to be an effective method for understanding the fish sex-determination system and its underlying mechanisms, as well as for contributing to the understanding of sex determination development in vertebrates.

Conclusions

In this study, three sex-specific markers for *M. amblycephala* were successfully identified through genome re-sequencing and biological information analysis. These markers can be rapidly and effectively used to identify the genetic sex of *M. amblycephala* using PCR amplification and agarose gel electrophoresis. Their specificity was confirmed in the breeding population. Additionally, we also successfully applied one of the markers to identify the genetic sex of *M. pellegrini*. In summary, this research is important for the ongoing study of the sex-determination mechanism and unisexual breeding of *M. amblycephala*. It also serves as a crucial resource and foundation for future sex identification and related studies of *Megalobrama* species.

Abbreviations

RAPD	Randomly amplified polymorphic DNA
AFLP	Amplified fragment length polymorphism
SSR	Simple sequence repeats
SNP	Single nucleotide polymorphisms
PCR	Polymerase chain reaction
YNL	Yuni Lake
PYL	Poyang Lake
LZL	Liangzi Lake
ZL	Hubei Zhenglong Aquatic Seed Industry Co., Ltd.
HH1	'Huahai No. 1'
PJ1	'Pujiang No. 1'
PJ2	'Pujiang No. 2'
GS	Guangshui
QJ	Qianjiang
EZ	Ezhou
HE	Hematoxylin and eosin
RAD-seq	Restriction site associated DNA sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10884-0.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors would like to thank all the staff of Huazhong Agricultural University who participated in data collection and sample analysis for the study.

Author contributions

YYF participated in the methodology, validation, formal analysis, investigation, writing-original draft, and visualization. LFL carried out the methodology, data curation, formal analysis, writing-review & editing, and funding acquisition. SLW, YY and YW carried out the methodology and data curation. ZXG participated in the conceptualization, writing–review & editing, supervision, project administration and funding acquisition. All authors read and approved the final manuscript.

Funding

This research was supported by the earmarked fund for China Agriculture Research System (CARS-45-01), the National Natural Science Foundation of China (32330108 and 32202929), the National Key Research and Development Program (2022YFD2400600/2022YFD2400604), the Major Science and Technology project of Hubei Province (2023BBA001), the Fund projects supporting high-quality seed industry development of Hubei Province (HBZY2023B009) and the Biological breeding project of Wuhan Municipal Bureau of Science and Technology (2022021302024854).

Data availability

Sequencing data were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under BioProject Accession Number PRJNA1095383. This includes sequence read archive (SRA) submissions for 257 individuals whole genome sequence reads (SRX24204205 to SRX24204208, SRX24204210 to SRX24204220, SRX24204223, SRX24204224, SRX24204228 to SRX24204246, SRX24204248 to SRX24204352, SRX24204354 to SRX24204376, SRX24204378 to SRX24204390, SRX24204392 to SRX24204403, SRX24204405 to SRX24204467, SRX24204470 to SRX24204474) used in this study. Other data generated or analyzed during this study are included in the main paper.

Declarations

Ethics approval and consent to participate

All the experimental procedures involved fish were approved by the Institutional Animal Care and Use Committee of Huazhong Agricultural University (Wuhan, China) (HZAUFI-2023-0043). All surgery was performed under MS-222 anesthesia, and all efforts were made to minimize suffering.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Received: 24 March 2024 / Accepted: 9 October 2024
Published online: 15 October 2024

References

1. Miya M, Gotoh RO, Sado T. MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries Sci. 2020;86:939–70.
2. Chen J, Hu W, Zhu ZY. Progress in studies of fish reproductive development regulation. Chin Sci Bull. 2013;58:7–16.
3. Li XY, Gui JF. Diverse and variable sex determination mechanisms in vertebrates. Sci China Life Sci. 2018;61:1503–14.
4. Kobayashi Y, Nagahama Y, Nakamura M. Diversity and plasticity of sex determination and differentiation in fishes. Sex Dev. 2013;7:115–25.
5. Mei J, Gui JF. Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. Sci China Life Sci. 2015;58:124–36.
6. Wang HP, Shen ZG. Sex Control in Aquaculture. Wiley-Blackwell. 2018;1–34.
7. Shen ZG, Wang HP. Molecular players involved in temperature-dependent sex determination and sex differentiation in teleost fish. Genet Sel Evol. 2014;46:26.
8. Gui JF, Zhu ZY. Molecular basis and genetic improvement of economically important traits in aquaculture animals. Chin Sci Bull. 2012;57:1751–60.
9. Chen JJ, Wang YL, Yue Y, Xia XH, Du QY, Chang ZJ. A novel male-specific DNA sequence in the common carp, *Cyprinus carpio*. Mol Cell Probe. 2009;23:235–9.
10. Zhi JH, Kang SL, Xia XH, Du QY, Chang ZJ. Identification of a novel male-specific DNA marker in loach (*Misgurnus anguillicaudatus*). J Mol Cell Biol. 2009;42:231–6.
11. Xia XH, Zhao J, Du QY, Zhi JH, Chang ZJ. Cloning and identification of a female-specific DNA marker in *Paramisgurnus dabryanus*. Fish Physiol Biochem. 2011;37:53–9.
12. Griffiths R, Orr KL, Adam A, Barber I. DNA sex identification in the three-spined stickleback. J Fish Biol. 2000;57:1331–4.
13. Wang D, Mao HL, Chen HX, Liu HQ, Gui JF. Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. Anim Genet. 2009;40:978–81.
14. Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR. A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics. 2001;158:727–34.
15. Lee BY, Penman DJ, Kocher TD. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. Anim Genet. 2003;34:379–83.
16. Woram RA, Gharbi K, Sakamoto T, Hoyheim B, Holm LE, Naish K, et al. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. Genome Res. 2003;13:272–80.
17. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, et al. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. P Natl Acad Sci Usa. 2012;109:2955–9.
18. Li MH, Sun YL, Zhao J, Shi HJ, Zeng S, Ye K, et al. A tandem duplicate of anti-nüllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. PLoS Genet. 2015;11:e1005678.
19. Lin AQ, Xiao SJ, Xu SB, Ye K, Lin XY, Sun S, et al. Identification of a male-specific DNA marker in the large yellow croaker (*Larimichthys crocea*). Aquaculture. 2017;480:116–22.
20. Meng XH, Dong LJ, Shi XL, Li XP, Sui J, Luo K, et al. Screening of the candidate genes related to low-temperature tolerance of *Fenneropenaeus chinensis* based on high-throughput transcriptome sequencing. PLoS ONE. 2019;14:e0211182.
21. Yadav CB, Bhareti P, Muthamilarasan M, Mukherjee M, Khan Y, Rathi P, et al. Genome-wide SNP identification and characterization in two soybean cultivars with contrasting Mungbean Yellow Mosaic India Virus disease resistance traits. PLoS ONE. 2015;10:e0123897.
22. Sun XY, Niu QH, Jiang J, Wang GF, Zhou P, Li J, et al. Identifying candidate genes for litter size and three morphological traits in Youzhou dark goats based on genome-wide SNP markers. Genes (Basel). 2023;14:1183.
23. Ou M, Yang C, Luo Q, Huang R, Zhang AD, Liao LJ, et al. An NGS-based approach for the identification of sex-specific markers in snakehead (*Channa argus*). Oncotarget. 2017;8:98733–44.
24. Zhang AD, Huang R, Chen LM, Xiong L, He LB, Li YM, et al. Computational identification of Y-linked markers and genes in the grass carp genome by using a pool-and-sequence method. Sci Rep. 2017;7:8213.
25. Liu HY, Pang MX, Yu XM, Zhou Y, Tong JG, Fu BD. Sex-specific markers developed by next-generation sequencing confirmed an XX/XY sex determination system in bighead carp (*Hypophthalmichthys nobilis*) and silver carp (*Hypophthalmichthys molitrix*). DNA Res. 2018;25:341.
26. Zhou YL, Wu JJ, Wang ZW, Li GH, Mei J, Zhou L, et al. Identification of sex-specific markers and heterogametic XX/XY sex determination system by 2b-RAD sequencing in redtail catfish (*Mystus Wyckioides*). Aquac Res. 2019;50:2251–66.
27. Suda A, Nishiki I, Iwasaki Y, Matsuura A, Akita T, Suzuki N, et al. Improvement of the Pacific bluefin tuna (*Thunnus orientalis*) reference genome and development of male-specific DNA markers. Sci Rep. 2019;9:14450.
28. Yang C, Huang R, Ou M, Gui B, Zhao J, He LB, et al. A rapid method of sex-specific marker discovery based on NGS and determination of the XX/XY sex-determination system in *Channa maculata*. Aquaculture. 2020;528:735499.
29. Han C, Zhu QY, Lu HM, Wang CW, Zhou XN, Peng C, et al. Screening and characterization of sex-specific markers developed by a simple NGS method in mandarin fish (*Siniperca chuatsi*). Aquaculture. 2020;527:735495.
30. Gao ZX, Luo WW, Liu H, Zeng C, Liu XL, Yi SK, et al. Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). PLoS ONE. 2012;7:e42637.
31. FAO. The State of World Fisheries and Aquaculture 2020. Sustainability in action. FAO, Rome, Italy: 2022.
32. Dong Q, Nie CH, Wu YM, Zhang DY, Wang XD, Tu T, et al. Generation of blunt snout bream without intermuscular bones by runx2b gene mutation. Aquaculture. 2023;567:739263.
33. Chen J, Liu H, Gooneratne R, Wang Y, Wang WM. Population genomics of *Megalobrama* provides insights into evolutionary history and dietary adaptation. Biology (Basel). 2022;11:186.
34. Ye XP, Cheng BL, Zhang TY. Analysis on the conservation and breeding of freshwater fish germplasm resources. Sci Fish Farm. 2006;11:6–7.
35. Li XY, Mei J, Ge CT, Liu XL, Gui JF. Sex determination mechanisms and sex control approaches in aquaculture animals. Sci China Life Sci. 2022;65:1091–122.
36. Tao WJ, Zhu X, Cao JM, Xiao HS, Dong JJ, Kocher TD, et al. Screening and characterization of sex-linked DNA markers in Mozambique tilapia (*Oreochromis mossambicus*). Aquaculture. 2022;557:738331.
37. Liu HY, Xia WW, Li BJ, Liu L, Wang YK, Luo Q, et al. Sex-specific markers developed by 2b-RAD and genome sequencing reveal an XX/XY sex-determination system in mud carp (*Cirrhinus molitorella*). Aquaculture. 2023;565:739131.
38. Zhang XH, Gao ZX, Luo W, Wen JF, Song W, Wang WM. Studies on morphological characteristics and genetic analysis of the gynogenesis blunt snout bream (*Megalobrama amblycephala*). Acta Hydrobiol Sin. 2015;39:126–32.
39. Wen M, Wang SY, Zhu CC, Zhang YX, Liu Z, Wu C, et al. Identification of sex locus and a male-specific marker in blunt-snout bream (*Megalobrama amblycephala*) using a whole genome resequencing method. Aquaculture. 2024;582:740559.
40. Bolger AM, Lohse M, Usadel B, Trimmomatic. A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
41. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–7.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
44. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
45. Yang C, Chen LM, Huang R, Gui B, Li YY, Li YY, et al. Screening of genes related to sex determination and differentiation in mandarin fish (*Siniperca chuatsi*). Int J Mol Sci. 2022;23:7692.
46. Hayes TB, Anderson LL, de Beasley Vr SR, Iguchi T, Ingraham H et al. Demasculinization and feminization of male gonads by atrazine: consistent effects across vertebrate classes. J Steroid Biochem Mol Biol. 2011;127.
47. Zhu CK, Liu HY, Pan ZJ, Cheng L, Sun YH, Wang H, et al. Insights into chromosomal evolution and sex determination of *Pseudobagrus ussuriensis* (Bagridae, Siluriformes) based on a chromosome-level genome. DNA Res. 2022;29:dsac028.

48. Wang T, Li Z, Yu ZX, Wang ZW, Lian ZQ, Du WX, et al. Production of YY males through self-fertilization of an occasional hermaphrodite in Lanzhou catfish (*Silurus lanzhouensis*). Aquaculture. 2021;539:736622.
49. Wen M, Feron R, Pan Q, Guguin J, Jouanno E, Herpin A et al. Sex chromosome and sex locus characterization in goldfish, *Carassius auratus* (Linnaeus, 1758). BMC Genomics. 2020;21:552.
50. Zhu CK, Liu HY, Cheng L, Pan ZJ, Chang GL, Wu N, et al. Identification of sex-specific sequences through 2b-RAD sequencing in *Pseudobagrus ussuriensis*. Aquaculture. 2021;539:736639.
51. Dai SM, Zhou YL, Guo XF, Liu M, Wang T, Li Z, et al. Sex-specific markers developed by genome-wide 2b-RAD sequencing confirm an XX/XY sex determination system in Chinese longsnout catfish (*Leiocassis longirostris*). Aquaculture. 2022;549:737730.
52. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.
53. Gui JF, Zhu ZY. Molecular basis and genetic improvement of important economic traits in aquatic animals. Chin Sci Bull. 2012;57(19):11.
54. Liu H, Guan B, Xu J, et al. Genetic manipulation of sex ratio for the large-scale breeding of YY super-male and XY all-male yellow catfish (*Pelteobagrus fulvidraco* (Richardson)). Mar Biotechnol (NY). 2013;15:321–8.
55. Liu S, Zheng J, Li F, Chi M, Cheng S, Jiang W, et al. Chromosome-scale assembly and quantitative trait locus mapping for major economic traits of the *Culter alburnus* genome using Illumina and PacBio sequencing with Hi-C mapping information. Front Genet. 2023;14:1072506.
56. Stöck M, Kratochvíl L, Kuhl H, Rovatsos M, Evans BJ, Suh A, et al. A brief review of vertebrate sex evolution with a pledge for integrative research: towards '*sexomics*'. Philos Trans R Soc Lond B Biol Sci. 2021;376(1832):20200426.
57. Jin L, Jia ST, Zhang W, Chen YB, Li SD, Liu P, et al. Identification of sex-specific DNA markers: providing molecular evidence for the ZW sex determination system in the redclaw crayfish (*Cherax quadricarinatus*). Aquaculture. 2022;546:737254.
58. Han C, Zhou XN, Lu HM, Zhu QY, Han LQ, Li SS, et al. A simple PCR-based genetic sex identification method in the blotched snakehead (*Channa maculata*) developed by high-throughput sequencing. Aquaculture. 2021;538:736579.
59. Star B, Tørresen OK, Nederbragt AJ, Jakobsen KS, Pampoulie C, Jentoft S. Genomic characterization of the Atlantic Cod sex-locus. Sci Rep. 2016;6:31235.
60. Xing TF, Li YL, Liu JX. Female-specific genomic regions and molecular sex identification of the clearhead icefish (*Protosalanx hyalocranius*). BMC Genomics. 2021;22:495.
61. He QW, Ye K, Han W, Yekefenhazi D, Sun S, Xu XD, et al. Mapping sex-determination region and screening DNA markers for genetic sex identification in largemouth bass (*Micropterus salmoides*). Aquaculture. 2022;559:738450.

Publisher's note