SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-023-2694-5

RESEARCH PAPER

Unveiling potential sex-determining genes and sex-specific markers in autotetraploid *Carassius auratus*

Kun Zhang[1], Xu Huang[1], Chongqing Wang[1], Xidan Xu[1], Xiaowei Xu[1], Xiaoping Dong[1], Qingwen Xiao[1], Jinhai Bai[1], Yue Zhou[1], Zhengkun Liu[1], Xinyi Deng[1], Yan Tang[1], Siyang Li[1], Enkui Hu[1], Wanjing Peng[1], Ling Xiong[1], Qinbo Qin[1,2,3*] & Shaojun Liu[1*]

[1]State Key Laboratory of Developmental Biology of Freshwater Fish, Engineering Research Center of Polyploid Fish Reproduction and Breeding of the State Education Ministry, College of Life Sciences, Hunan Normal University, Changsha 410081, China;
[2]Nansha-South China Agricultural University Fishery Research Institute, Guangzhou 511457, China;
[3]Hunan Yuelu Mountain Science and Technology Co., Ltd., for Aquatic Breeding, Changsha 410081, China

*Corresponding authors (Qinbo Qin, email: qqb@hunnu.edu.cn; Shaojun Liu, email: lsj@hunnu.edu.cn)

Received 11 March 2024; Accepted 28 July 2024; Published online 9 August 2024

Autotetraploid *Carassius auratus* is a stable hereditary autotetraploid fish resulting from the hybridization of *Carassius auratus* red var. (RCC, ♀) × *Megalobrama amblycephala* (BSB, ♂), containing four sets of RCC chromosomes. However, the molecular mechanism underlying the determination of sex in this species remains largely unknown. Currently, there lacks a full understanding of the molecular mechanisms governing sex determination and specific molecular markers to differentiate sex in this species. In this study, 25,801,677 SNPs (Single-nucleotide polymorphism) and 6,210,306 Indels (insertion-deletion) were obtained from whole-genome resequencing of 100 individuals (including 50 female and 50 male). Further identification confirmed the candidate chromosomes as Chr46B, with the sex-determining region located at Chr46B: 22,500,000–22,800,000 bp. Based on the male-specific insertion (26 bp) within the candidate sex-determining region, a pair of sex-specific molecular markers has been identified. In addition, based on the screening of candidate sex-determining region genes and RT-qPCR validation analysis, *ADAM10*, *AQP9 and tc1a* were identified as candidate sex-determining genes. These findings provide a robust foundation for investigating sex determination mechanisms in fish, the evolution of sex chromosomes, and the development of monosex populations.

autotetraploid *Carassius auratus* | sex-specific marker | sex determination gene | *Re*-sequencing

INTRODUCTION

Sex determination (SD) is the biological mechanism responsible for establishing the course of sexual development, and in aquatic animals, sex is typically a significant attribute (Weng et al., 2022). Aquaculture animals exhibit a remarkably diverse range of sexual phenotypes and underlying mechanisms, making them ideal for conducting sex-determination studies that are essential to the life sciences (Li et al., 2022; Mei and Gui, 2015). Improved understanding of the molecular mechanisms of sex determination can facilitate the application of reproductive and sex control techniques in aquaculture practices (Devlin and Nagahama, 2002). Fish, as lower vertebrates, have sex chromosomes that are at a primitive stage of evolution, leading to highly diverse and complex sex determination systems (Li and Gui, 2018; Li et al., 2022). Furthermore, fish models play a crucial role in advancing our understanding of vertebrate sex determination mechanisms and evolutionary history (Zhang et al., 2022). Therefore, studying fish sex determination offers valuable opportunities for gaining a deeper understanding about the origin and evolution of vertebrate sex chromosomes (Ellegren, 2011; Mei and Gui, 2015; Zhang et al., 2022). Most aquaculture fish species display sexual dimorphism, encompassing differences in growth, body size, and other economically significant characteristics (Leinonen et al., 2011; Li et al., 2022; Liang et al., 2020). Therefore, cultivating rapidly and robustly growing mono-sex populations allows for better management and enhanced aquaculture profitability (Wang et al., 2009). In many fish species, sexual chromosome differentiation is minimal, which hinders the distinction of sexes based on the morphology of the sex chromosomes (Nagahama et al., 2020). Additionally, environmental factors such as pH, temperature, density, and social interactions can easily influence the sex of fish, resulting in differences between their genetic sex (genotype) and physiological sex (ovary or testis) (Kobayashi et al., 2012; Reddon and Hurd, 2013; Sandra and Norma, 2010; Shen and Wang, 2014; Yamamoto et al., 2019). Sex-specific DNA markers, which are closely linked to sex, are essential tools for studying sex determination mechanisms and facilitating marker-assisted sex control (Chen et al., 2008). Traditionally, sex-specific DNA markers have been developed in some aquatic species with molecular methods, including random amplified polymorphic DNA (RAPD) (Vale et al., 2014), simple sequence repeat (SSR) (Chen et al., 2012; Shikano et al., 2011), and amplified restriction fragment polymorphism (AFLP) (Ezaz et al., 2004; Lee et al., 2011). Whole-genome resequencing allows for rapid and efficient analysis of base differences between males and females, facilitating the development of more accurate sex-specific molecular markers. Currently, some sex-specific molecular markers developed based on SNPs have been applied to *Larimichthys crocea* (Lin et al., 2018), *Mystus wyckioides* (Zhou et al., 2019), *Pseudobagrus ussuriensis* (Zhu et al., 2021), *Cynoglossus semilaevis* (Zhang et al., 2019), *Litopenaeus vannamei* (Perez-

Enriquez et al., 2020) and so on (Gammerdinger et al., 2016; Gao et al., 2020; Kamiya et al., 2012; Shi et al., 2018; Zhang et al., 2022). By using insertions/deletions as sex-specific molecular markers, the genetic sex of samples can be determined through PCR and agarose gel electrophoresis (Chen et al., 2022; Lin et al., 2017; Weng et al., 2022; Xiao et al., 2020; Zhang et al., 2017; Zheng et al., 2020). Moreover, with the rapid progress in genome sequencing technology and comparative genomics, it has become increasingly feasible to delineate sex determination regions in non-model species with complex genomes using genomic data (Bachtrog et al., 2014). Furthermore, this presents an opportunity to develop molecular markers for sex determination and sex-specific markers in autotetraploid *Carassius auratus*.

Polyploidy played a significant role in the evolutionary history of vertebrates and other eukaryotic organisms (Wood et al., 2009). Polyploidy occurs when an uncommon event during mitosis or meiosis results in the formation of gametes with multiple sets of chromosomes (Otto, 2007). In our previous study, the world's first fertile autotetraploid *Carassius auratus* (4*n*=200, RRRR, abbreviated as 4nRR) was produced through distant hybridization from distant hybridization *Carassius auratus* red var. (2*n*=100, abbreviated as RCC) (♀) X *Megalobrama amblycephala* (2*n*=48, abbreviated as BSB) (♂) (Qin et al., 2014). The 4nRR originates from a whole-genome duplication event of the RCC, leading to the production of diploid eggs due to the pairing of diploid-like chromosomes during female meiosis (Wang et al., 2021). Thus, compared to diploid fish, developing sex-specific molecular markers for 4nRR will be more challenging. A variety of autotriploid or allotriploid fish with excellent traits can be prepared using interspecific crosses between 4nRR lines and RCC or other diploid fish. Typically, it is necessary to dissect fish and conduct histological observations to further determine their sex. The same procedures are required for sex identification in 4nRR. Moreover, progeny cross experiments necessitate extensive random detection, which requires a lot of time and effort (Liu et al., 2023). Therefore, there is an urgent need for a direct, reliable, and non-invasive method to determine gender, facilitating the breeding management of 4nRR. The development of sex-specific molecular markers and the identification of sex determination have been facilitated by a chromosome-level genome assembly of 4nRR (The genome of 4nRR is unpublished). Therefore, developing a new genetic technique to screen sex-specific DNA sequences in 4nRR is imperative, despite the challenges posed by polyploid fish, as no sex-specific genetic markers have been reported to date.

In this study, genome re-sequencing was applied to investigate sex determination and genetic variations (SNPs and Indels) in 4nRR, aiming to effectively identify sex-specific molecular markers and sex determination genes. In addition, based on genome resequencing, the sex-determining region and candidate genes for sex determination in 4nRR were identified. A rapid and reliable molecular method based on specific insertion/deletion sequences has been developed to identify the genetic sex of individuals. Furthermore, the developed Indel molecular markers were validated in many male and female individuals of 4nRR using PCR. These research findings contribute to the controlled breeding of 4nRR, enhancing breeding efficiency and providing insights into the sex determination mechanism of 4nRR.

RESULTS

Re-sequencing data and SNP calling

A total of 5784.76 Gb of raw data was obtained through whole-genome resequencing of both male and female populations. Among them, the size of raw data for the female and male populations was 2943.94 and 2840.82 Gb respectively (Table 1, Table S1 in Supporting Information). After filtering the raw data, we obtained a total of 5633.74 Gb of clean data, with 2866.81 Gb for the female population and 2766.93 Gb for the male

Table 1. Statistical results of quality control information of sequencing data

Sample	Raw base (Gb)	Clean base (Gb)	Clean reads Q20	Clean reads Q30	Clean rate (%)	GC content (%)
FJF1	58.78	58.57	97.11	90.30	99.81	38.17
FJF2	68.00	67.78	97.13	90.43	99.83	37.95
FJF3	67.33	67.10	97.56	91.99	99.82	38.11
FJF4	65.38	65.10	96.82	89.36	99.78	38.17
FJF5	71.79	71.63	97.51	91.63	99.87	37.80
FJF6	67.39	67.13	97.10	90.26	99.80	38.02
FJF7	66.97	66.69	97.80	93.49	99.80	38.05
FJF8	65.42	65.11	97.74	93.20	99.77	38.38
FJF9	70.67	70.30	95.96	88.59	99.70	37.95
FJF10	64.53	64.31	98.11	94.31	99.83	38.08
MJF1	71.62	71.30	97.80	93.48	99.79	38.37
MJF2	66.34	66.08	97.85	93.58	99.81	38.09
MJF3	69.06	68.77	97.86	93.60	99.80	38.10
MJF4	67.09	66.81	97.84	93.33	99.78	37.84
MJF5	66.63	66.35	97.96	93.73	99.77	37.52
MJF6	69.82	69.53	98.12	94.18	99.77	38.28
MJF7	67.35	67.07	98.02	93.91	99.77	38.08
MJF8	67.43	67.14	97.80	93.27	99.76	37.90
MJF9	66.05	65.77	97.84	93.36	99.76	38.31
MJF10	71.06	70.75	97.93	93.62	99.76	38.09

https://doi.org/10.1007/s11427-023-2694-5

population (Table 1, Table S1 in Supporting Information). The proportion of clean reads with quality scores greater than or equal to Q30 ranges from 88.59% to 96.46% and the proportion of data GC was below 39%, indicating that the sequencing results have good quality (Table 1, Table S1 in Supporting Information). The size range of the mapping rate of the filtered clean data to the reference genome by BWA is 99.79% to 99.96% (Table 2, Table S2 in Supporting Information). Among them, the proportion of reads correctly aligned (with suitable insertion fragments and directions) to the reference genome ranges from 87.52% to 98.39% (Table 2, Table S2 in Supporting Information). After alignment, the average coverage depth of the reference genome per sample ranges from 11.34× to 22.64× (Table 2, Table S2 in Supporting Information).

Using the aligned results, SNP calling was performed using GATK, resulting in a total of 45,806,409 SNPs and 12,838,623 Indels. To enhance the quality of the SNPs, additional filtering was applied to the extracted SNPs for subsequent GWAS analysis. Finally, after further filtering, a total of 25,801,677 SNPs and 6,210,306 Indels were obtained. The distribution of SNPs and Indels on each chromosome was shown in SNP and Indel density map (Figure 1).

Genetic structure of population

Principal component analysis (PCA) can reveal information about group clustering and outlier samples. Based on the SNPs after screening, PCA analysis of two populations was performed by using the VCF2PCACluster. After obtaining the PCA results, the analyzed samples were exhibited using the values of the first three PCs. The two-dimensional plots of the first three PCA are presented in Figure 2A and B; however, genetic clustering between male and female populations is not evident. Furthermore, genetic distances between populations were computed using the VCF2Dis software, and an ML tree was constructed based on these genetic distances. The results illustrated in Figure 2C also fail to differentiate genetic clustering between the two populations.

Identification and location of sex-linked variants

In population genetics, the F_{st} index is widely recognized as a standard measure for quantifying the degree of differentiation among populations. It is generally considered that F_{st} with greater than 0.15 indicates a significant genetic differentiation between two populations. Therefore, the higher the F_{st} value, the greater the degree of genetic differentiation between different populations, and the differences are more pronounced. The vcftools software was used to calculate the F_{st} between female and male groups, the results were plotted across the entire genome as shown in the Figure 3A. The results indicated that there are F_{st} values greater than 0.25 on chromosomes 19A, 31A, 13B, 17B, 28B, 39B and 46B. Further analysis using ED values to assess the frequency differences of SNPs between the two populations, with the results shown in the Figure 3B. Then, the distribution of the number of differentiating SNPs within intervals across the genome was calculated, as shown in the Figure 3C. Combining the results from the three methods mentioned above, Chr46B were identified as candidate chromosomes for 4nRR. Further, the regions Chr46B:22,500,000–22,800,000 bp are identified as

Table 2. Statistics of quality-controlled data alignment to the reference genome[a)]

Sample	MapRate (%)	MapPERate (%)	Properly rate (%)	Coverage (%)	MeanDepth (*X*)	Coverage(%)		
						Depth 1×	Depth 5×	Depth 10×
FJF1	99.95	99.92	97.54	100.00	18.50	96.21	92.51	88.06
FJF2	99.95	99.91	96.69	100.00	21.37	96.77	93.73	87.71
FJF3	99.95	99.92	97.53	100.00	21.19	96.76	93.80	88.52
FJF4	99.93	99.89	97.56	100.00	20.55	97.02	93.84	87.48
FJF5	99.96	99.94	98.39	100.00	22.64	96.38	93.62	86.58
FJF6	99.95	99.92	97.47	100.00	21.21	96.37	92.96	88.60
FJF7	99.93	99.88	97.53	100.00	21.04	96.66	93.63	87.99
FJF8	99.92	99.86	96.85	100.00	20.53	96.43	92.67	89.45
FJF9	99.92	99.88	98.06	100.00	22.17	96.69	93.46	88.55
FJF10	99.93	99.88	97.47	100.00	20.29	96.80	93.72	85.17
MJF1	99.92	99.87	96.92	100.00	22.48	96.95	93.88	88.57
MJF2	99.93	99.88	97.33	100.00	20.85	96.54	93.38	87.74
MJF3	99.92	99.86	96.95	100.00	21.68	97.00	93.96	88.26
MJF4	99.92	99.85	97.48	100.00	21.09	96.22	93.10	88.83
MJF5	99.92	99.86	97.26	100.00	20.93	96.84	93.89	88.04
MJF6	99.91	99.84	96.55	100.00	21.93	96.53	93.21	88.52
MJF7	99.92	99.86	96.91	100.00	21.16	96.82	93.79	87.39
MJF8	99.90	99.83	96.61	100.00	21.16	96.87	93.85	88.62
MJF9	99.92	99.85	96.67	100.00	20.75	96.80	93.39	87.39
MJF10	99.92	99.85	97.21	100.00	22.32	96.98	94.18	89.18

a) MapRate (%): All reads mapped to the reference genome sequence ratio; MapPERate (%): The percentage of reads that align in pairs to the reference genome sequence; Properly rate (%): The percentage of reads correctly aligned (with appropriate insert size, orientation, etc.) to the reference genome sequence; Coverage (%): Coverage of the reference genome sequence in sequencing data; MeanDepth (*X*): Average coverage depth of the reference sequence; Depth1×/5×/10× Coverage(%): The coverage rate of reads with coverage depths greater than or equal to 1×/5×/10×, respectively.

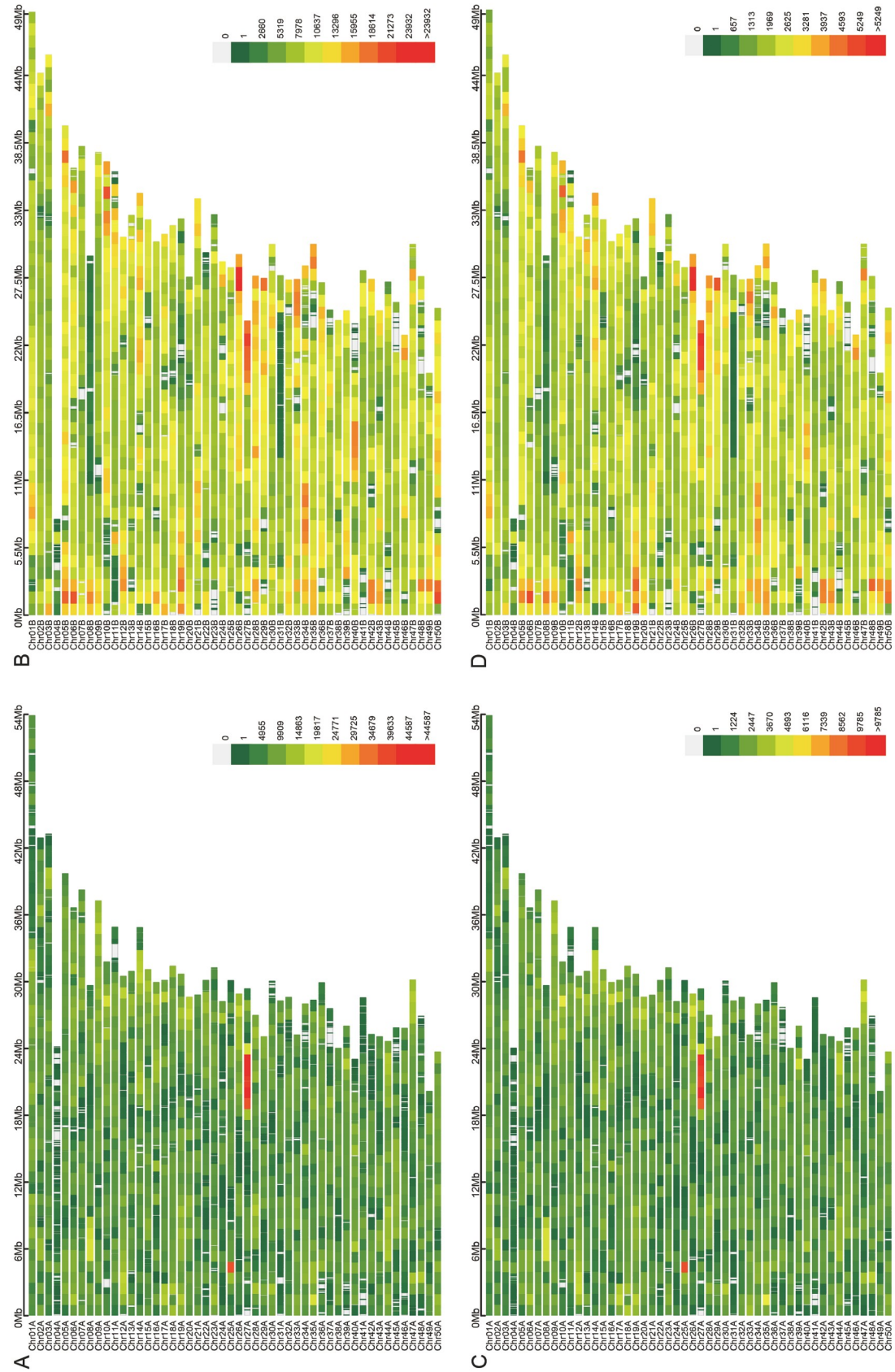

Figure 1. The hot map of SNP density and Indel on each chromosome A (ChrA) and chromosome B (ChrB) in autotetraploid *Carassius auratus*. A, SNP density heatmap located on ChrA; B, SNP density heatmap located on ChrB; C, Indel density heatmap located on ChrA; D, Indel density heatmap located on ChrB.

https://doi.org/10.1007/s11427-023-2694-5

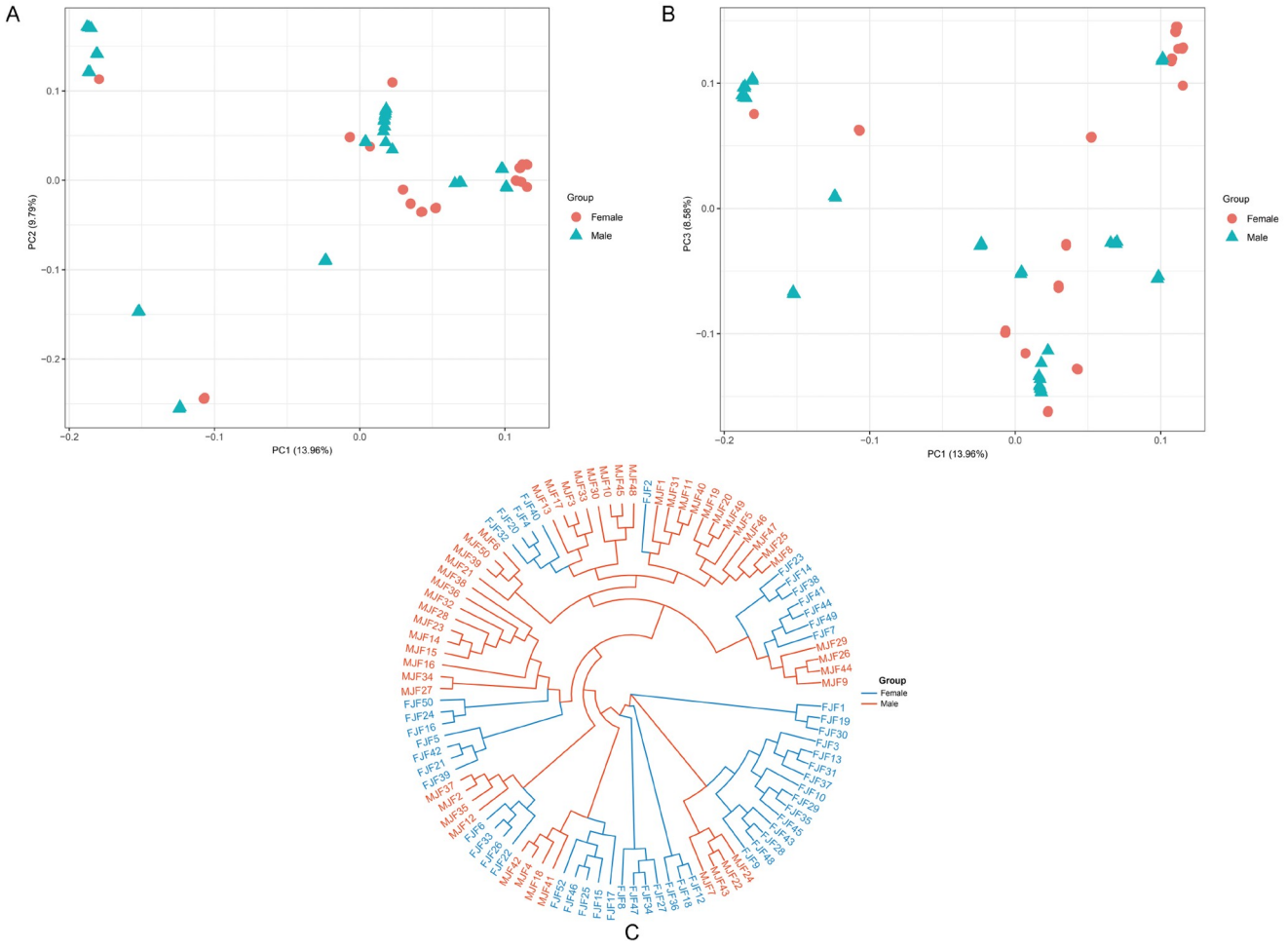

Figure 2. The relationship between PCA (Principal Component Analysis) analysis results and ML (Maximum Likelihood) phylogenetic trees based on SNP from male and female populations. A, The scatter plot of PC1 against PC2 in PCA. B, The scatter plot of PC1 against PC3 in PCA. Note: Circles represent sample group female, triangles represent sample group male. C, The ML tree based on SNP data from male and female populations, note: Highlighted in red font are the male population, and in blue font are the female population.

candidate regions for sex determination.

Application of the 4nRR sex detection PCR method

The re-sequencing alignment results were visualized with the software of IGV. Subsequently, based on the identified candidate chromosomes and the sex-determining region, we manually examined the variations using the IGV software. As shown in Figure 4, there was a 26 bp Indel (GGTTCAATAATGACTGATCACAGGTT) sequence in all sample of male that was found between 22,647,031 and 22,647,032 bp in genome of Chr46B. We will used the faidx function of samtools to extract the fragments in the upper and lower regions of the Indel, with the extraction range being "Chr46B:22,646,967–22,647,223". Designing primer pairs using Primer Premier v6.0 software, where the forward primer was 5′-GCTTACGGTTCAATAATGAC-3′, and the reverse primer was 5′-TCCCTAAGCTCTATTGTTCTC-3′ (Figure 5A). The primer was chosen to rapidly build a PCR-based method for identification of the genetic sex of 4nRR. Using the primer pair, PCR was performed, followed by 2% agarose gel electrophoresis. As shown in the Figure 5B and Figure S1, a single band was observed in all male individuals, while no bands were detected in the female individuals. For both male and female individuals, these primers can be used to differentiate their genetic sex.

Functional annotation of candidate genes by GO and KEGG pathway analysis

To further investigate the linkage disequilibrium (LD) among loci in the sex-determining region, LD analysis was performed on the SNPs in the genomic region where these six candidate genes were located. These six candidate genes are located on Chr46B: 22,500,000–22,800,000, and the SNP locus in these regions are in linkage disequilibrium (Figure 6). We found that most SNPs in this region are in a high level of linkage disequilibrium, with F_{st} than 0.25. To further explore the main biochemical and signaling pathways involved by candidate genes, we conducted pathway enrichment analysis based on the GO and KEGG databases. After comparing with the NR protein database, six genes were further chosen for GO and KEGG functional enrichment analysis (Table S3 in Supporting Information). As displayed in the Figure 7A, a total of five different genes (*ADAM10*, *AQP9*, *pclaf*, *RPS17*, and *tc1a*) were enriched in

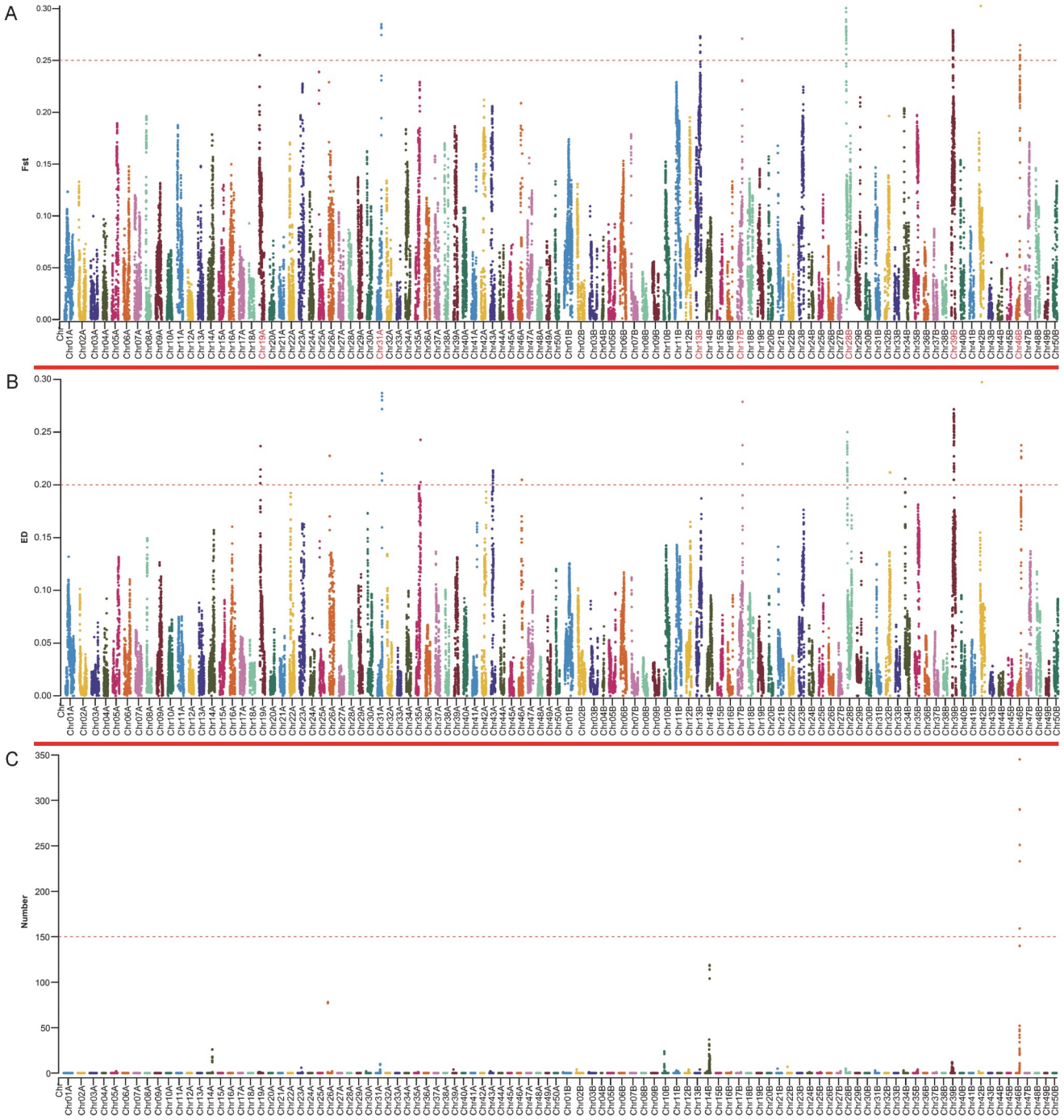

Figure 3. The Manhattan plot showed the sex significantly associated loci identified by GWAS. A, The *x*-axis indicates the chromosome numbers that were labeled with different colors. The *y*-axis is the F_{st} for each SNP. The horizontal dashed line is the position where the corresponding threshold F_{st} value is 0.25. B, The *x*-axis indicates the chromosome numbers that were labeled with different colors. The *y*-axis represents the Euclidean distance (ED) of SNP frequencies between male and female populations. C, The *x*-axis indicates the chromosome numbers that were labeled with different colors. The *y*-axis represents the total number of SNPs with abs (ΔSNP frequency) greater than 0.5 between groups. The window size for analyzing all SNPs is 200 kb.

various GO terms (Table S4 in Supporting Information). We conducted enrichment analysis using KEGG, and selected all enriched pathways to be displayed in a bubble chart (Figure 7B). In this case, only three genes (*ADAM10*, *AQP9*, and *RPS17*) were enriched in different pathways among the three KEGG enrichment pathways (Table S5 in Supporting Information).

Expression of candidate genes in the gonads

To validate the differential expression of these candidate genes in male and female gonadal tissues, nine genes were selected based on their specificity for further validation through RT-qPCR (Real-time quantitative PCR). The results indicated that the expression levels of three genes (*ADAM10*, *AQP9*, and *tc1a*) were

https://doi.org/10.1007/s11427-023-2694-5

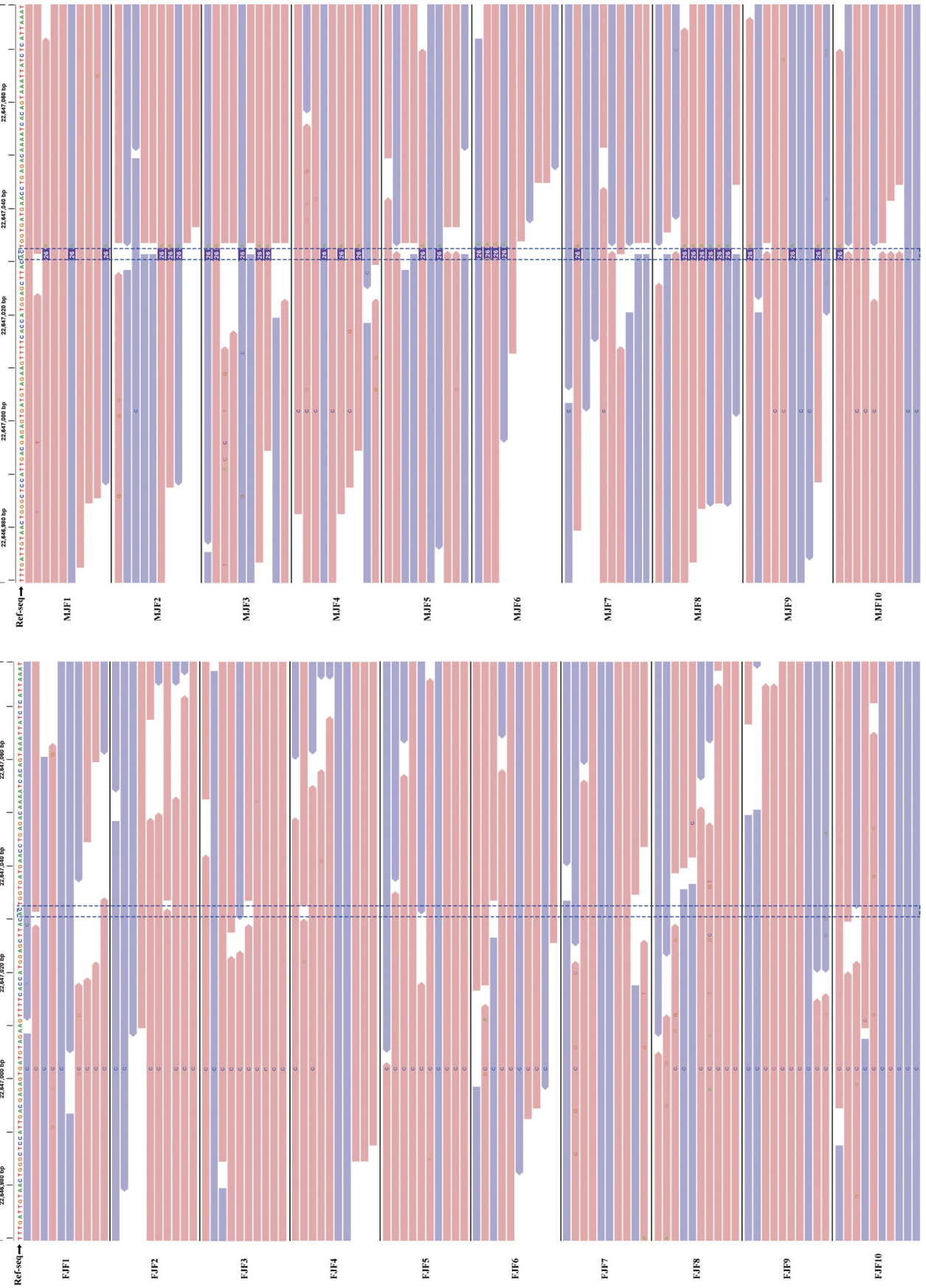

Figure 4. Comparative alignment of the sequencing reads from both sexes. The location within the dashed box represents a male-specific insertion fragment (26 bp), which is situated at Chr46B:22,647,031–22,647,032 bp.

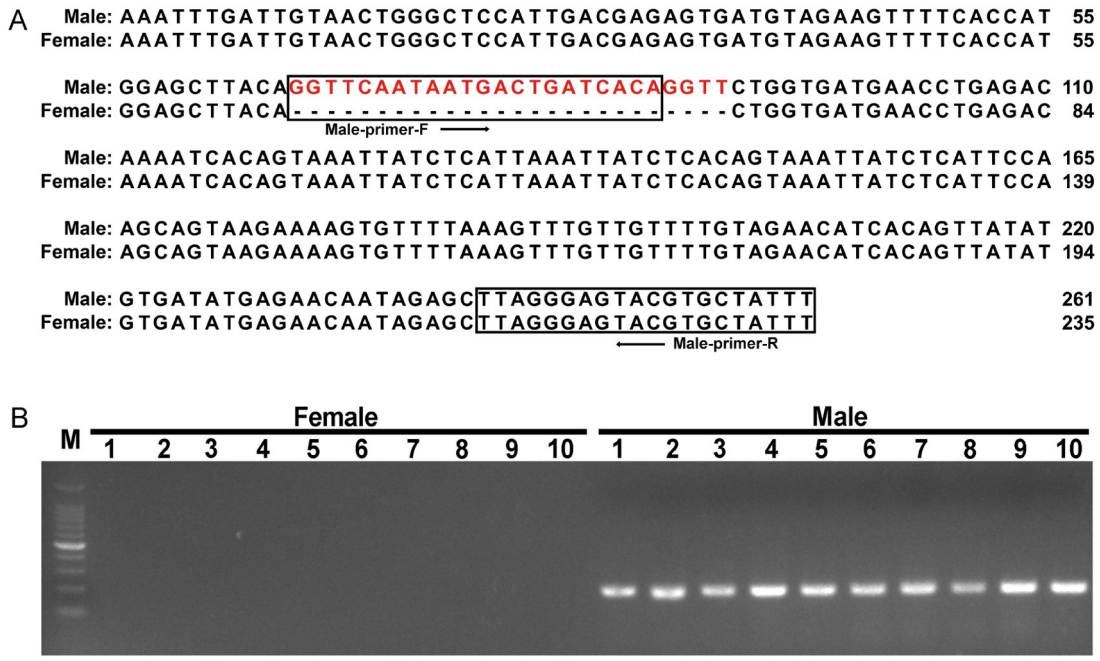

Figure 5. Primer amplification sequences and their specific molecular marker validation. A, Male-specific sequence and primer design, with red font indicating the 26 bp male-specific insertion sequence, and the black box highlighting the primer design sequence. B, Application of the PCR-based genetic sex identification method in autotetraploid *Carassius auratus* (female and male) population.

consistently higher in testis tissues across different stages compared to ovary tissues ($P<0.05$) (Figure 8A, 8B, and 8F). Additionally, the expression levels of the *pclaf* gene were consistently higher in ovary tissues across different stages compared to testis tissues ($P<0.0001$) (Figure 8D). *LOC107574091* exhibited varying expression patterns in different gonadal tissues across developmental stages; at 3 months of age, there was no significant difference in its expression levels between the gonadal tissues; at 5 months, its expression was significantly higher in the testis compared to the ovary ($P<0.05$); at 8 and 12 months, the expression levels were significantly higher in the ovary compared to the testis ($P<0.0001$) (Figure 8C). Furthermore, at 3 and 5 months of age, the expression levels of the RPS17 gene were higher in testis tissues compared to ovary tissues ($P<0.001$). At 8 and 12 months of age, however, the expression levels of this gene were higher in ovary tissues compared to testis tissues ($P<0.001$) (Figure 8E). While, we speculated that the differential expression observed in various gonadal tissues may be associated with the mechanisms underlying sex determination.

DISCUSSION

Compared to higher vertebrates, plants and insects, fish have more chromosomes and smaller chromosomes and morphology, making fish karyotype analysis become challenging. Therefore, identifying the sex of fish through chromosomal karyotyping poses significant challenges. As in other fishes, the sex of 4nRR is difficult to distinguish by chromosome karyotype. Additionally, previous studies have revealed that the 4nRR genome originates from a genome-wide replication of the RCC, thereby raising the difficulty in developing sex-specific molecular markers for 4nRR. Traditional techniques such as RFLP, AFLP, SSR, and RAPD have been successfully employed in the development of sex-specific markers in many fish species (Devlin et al., 1991; Griffiths et al., 2000; Kovács et al., 2000; Shikano et al., 2011). However, these methods require meticulous primer screening, and most of the obtained markers are too short for effective primer design (Sun et al., 2014). In recent years, advancements in sequencing technologies have enabled second-generation high-throughput sequencing methods to readily identify distinctions between X and Y (or Z and W) chromosomes at low levels of differentiation. Many sex-specific molecular markers have now been successfully developed in some aquatic organisms based on high-throughput sequencing. Whole-genome resequencing technology is a more efficient genotyping strategy, especially in the detection of large sex-linked regions. It can identify numerous polymorphic sites distributed across the entire genome, as observed in species such as *Channa maculate* (Han et al., 2021), *Micropterus salmoides L.* (Du et al., 2021), *Scatophagus tetracanthus* (Peng et al., 2023), and so on (MascaLi et al., 2022). However, these fish species are all diploid organisms. In addition, the same methods have also been applied to the development of sex-specific molecular markers in the octoploid sturgeon (Ruan et al., 2021). Therefore, these results indicated that whole-genome resequencing for gene genotyping exhibits sufficient efficiency and accuracy, making it applicable for the identification of sex-specific markers. Here, for the first time, we utilized whole-genome re-sequencing to identify efficient and accurate sex-specific markers in 4nRR. Interestingly, the forward primer for this male-specific molecular marker was precisely designed within the 26 bp sequence unique to males, and this male-specific sequence was found to exist in all male individuals. Additionally, by comparing and annotating the reference genome of the female 4nRR, we found that the inserted 26 bp sequence unique to males is in a non-coding sequence. However, whether the 26 bp sequence unique to the male-specific insertion will affect the expression of genes upstream and downstream of the male genome requires further exploration.

https://doi.org/10.1007/s11427-023-2694-5

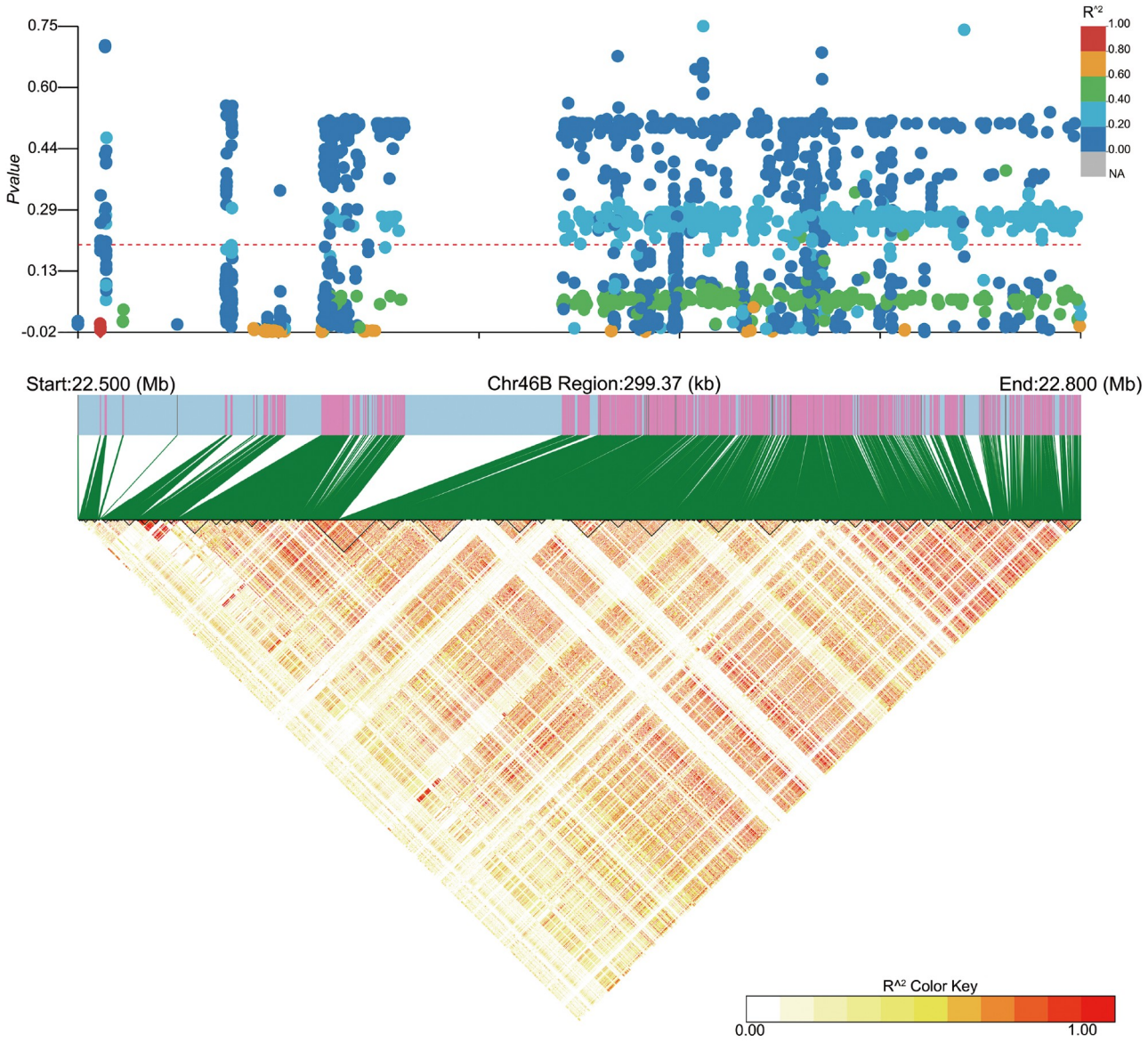

Figure 6. Linkage Disequilibrium (LD) analysis of Chromosomal positions of candidate sex determination genes. LD analysis spanning the physical position from 22.500 to 22.800 Mb of chromosome 46B. Top shows the *P*-values of SNPs in the genomic region of candidate sex determination genes, while bottom displays the LD heatmap of chromosomal positions of candidate sex determination genes.

The 4nRR, as an autotetraploid fish, can produce triploid fish when crossed with other diploid fish. These triploids may exhibit advantages in traits such as growth rate, disease resistance, or other economic characteristics, which are important for fish farming and enhance economic value. Nevertheless, significant challenges persist in determining the genetic sex of individuals, especially at an early stage, which hampers the successful implementation of sex-controlled breeding and the understanding of sex determination mechanisms in this species. In this study, we performed whole-genome resequencing and conducted genome-wide association studies (GWAS) analysis on ten individuals each of female and male 4nRR. In this study, a total of 25,801,677 SNPs (Figure 1A and B) and 6,210,306 Indels were identified (Figure 1C and D). Subsequently, candidate sex chromosomes and sex determination regions for 4nRR were determined based on SNPs using three methods: F_{st}, ED, and ΔSNP frequency. These identified candidate sex chromosomes for 4nRR were Chr46B, and the sex determination regions were determined as follows: Chr46B: 22,500,000–22,800,000 bp. Based on the candidate sex determination regions and identified sexually dimorphic SNP and Indel analyses, numerous sexually dimorphic variations were screened, including both SNP and Indel. Finally, a male-specific 26 bp insertion sequence was identified between 22,647,031 and 22,647,032 bp on Chr46B (Figure 4). Based on this male-specific insertion sequence, a pair of sex-specific primers was designed, which can only amplify a single specific band in male individuals (Figure 5B, Figure S1 in Supporting Information). We successfully developed, for the first time, sex-specific molecular markers for 4nRR, providing essential molecular tools for assisted sex control breeding, as well as elucidating the molecular mechanisms of sex determination and differentiation.

Genetic sex determination constitutes a fundamental biological process and is widespread among vertebrates. The mechanisms

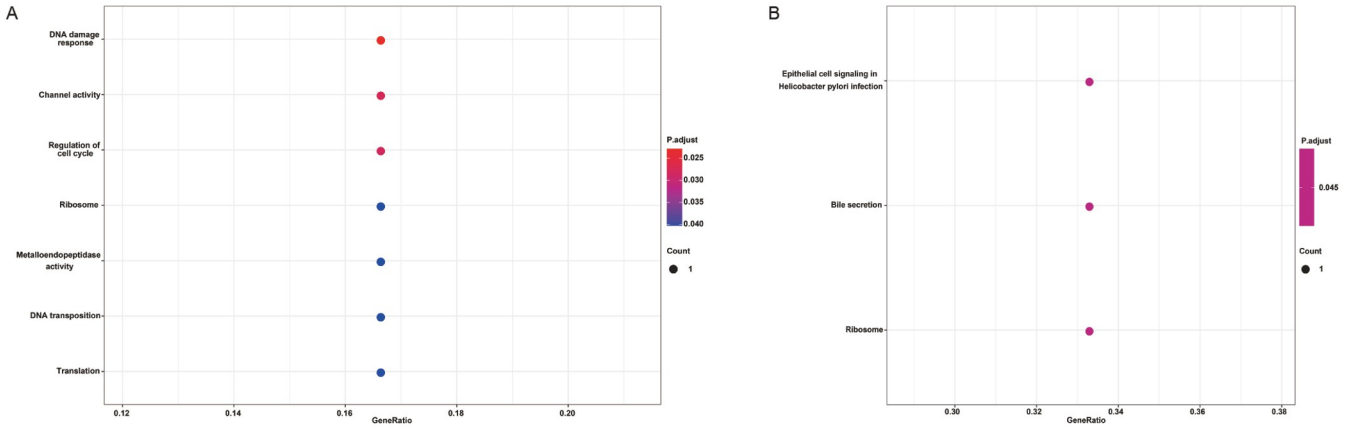

Figure 7. Functional enrichment bubble plot of candidate sex determination genes. A, Gene Ontology functional enrichment bubble plot of candidate sex determination genes. B, An overview of the KEGG pathways enriched in candidate sex determination genes. The *y*-axis represents the enriched pathways for candidate sex determination genes, while the *x*-axis indicates the gene ratio.

governing sex determination exhibit diversity across various taxonomic groups, especially in teleost fishes. However, the candidate genes and specific mechanisms underlying sex determination remain unclear for most teleost fishes. A key sex-determining gene in a species must meet three essential criteria: firstly, it must reside within the sex determining region and exhibit a close linkage to the sex locus; secondly, it should demonstrate distinct expression patterns during the critical window of sex differentiation; and lastly, it must be indispensable and fully capable of initiating testis development (Yano et al., 2012). In this study, through the method of GWAS analysis based on F_{st}, ED, and ΔSNP frequency, some genes associated with sex selection have been identified. We identified six candidate genes in the sex-determining candidate region, suggesting that these genes may be associated with the sex determination in 4nRR. GO and KEGG enrichment analyses give insights into the functions and roles of genes, as well as their involvement in specific biological processes, pathways, and diseases. These tools are crucial for interpreting biological data and guiding future research. In this study, GO and KEGG functional enrichment analyses were conducted on the candidate genes. It was observed that five genes (*tc1a*, *ADAM10*, *AQP9*, *pclaf*, and *RPS17*) were enriched in various GO terms, and three genes (*ADAM10*, *AQP9*, and *RPS17*) were enriched in different KEGG pathways. The genes *ADAM10*, *AQP9*, and *RPS17* were simultaneously enriched in various GO terms and KEGG pathways. ADAMs (disintegrins and metalloproteinases) family is crucial for sperm and egg fusion, development, inflammation, cellular adhesion, and migration (Blobel, 1997; Blobel, 2005; Huovila et al., 1996; Seals and Courtneidge, 2003). In this study, we found that *ADAM10* exhibited higher expression levels in testis compared to ovary tissues at various stages of gonadal development, with the highest expression observed in 5-month-old gonadal tissues compared to other ages (Figure 8A). Additionally, studies have shown high expression of *ADAM10* in the gonadal tissues of *Eriocheir sinensis*, suggesting that this gene may be involved in the reproductive development of its gonadal tissues (Li et al., 2015). Therefore, based on the expression levels of *ADAM10* in the gonads, we speculate that this gene plays an important role in the developmental process of 4nRR testis. Aquaporins (AQPs) are widely recognized as water channels that function as integral membrane pore proteins across all biological membranes, crucially facilitating water transport between cells (Monsang et al., 2019). Research indicates that in females, AQPs play crucial roles in fertilization, embryo formation, and implantation. Similarly, in males, AQPs are involved in sperm maturation, storage, transport, and release, contributing to increased sperm concentration (Huang et al., 2006). The *AQP7* and *AQP8* genes within the aquaporin gene family are observed to appear in the testes of mice as reproductive cells mature, and they participate in the process of spermatogenesis (Kageyama et al., 2001). In this study, *AQP9* showed consistently higher expression levels in testis tissues compared to ovary tissues across different developmental stages, with an increase observed as the gonads mature (Figure 8B). Furthermore, we found that *AQP9* exhibits expression patterns like *AQP3a* within the AQP gene family in *Cyprinus carpio*, with significantly higher expression levels observed in the testis compared to the ovary (Monsang et al., 2019). As a gene within the ribosomal proteins (RPs), *RPS17* has been found to exhibit changes at the cellular level in the transcription of RP transcripts in response to variations in growth and development (Kuang et al., 2020). In the zebrafish, the expression of RP genes was observed to undergo a continuous, coordinated increase from the initiation of mid-blastula transition until hatching in the model (Small et al., 2009). Some researchers discovered that in *Nile Tilapia*, the quantity and expression levels of differentially expressed RP genes in the XY gonads were higher than those in the XX gonads at 90 and 180 dah (Kuang et al., 2020). In this study, it was observed that the expression levels of the RPS17 gene gradually decreased across different ages of testis tissues, whereas in ovary tissues, the expression levels were lowest at 3 months and highest at 5 months of age. Additionally, at 8 and 12 months, the expression levels in the ovary were consistently higher than those in the testis (Figure 8E). Therefore, we speculate that the *RPS17* may be involved in the development of the 4nRR gonads. The PCNA-associated factor (*pclaf*, also known as *PAF15*) gene is involved in several critical cellular processes and plays a role in DNA repair during replication by interacting with PCNA (proliferating cell nuclear antigen) (Simpson et al., 2006). The expression level of the *pclaf* gene in ovarian tissues is higher than in testicular tissues across different ages, and it gradually increases with age in the ovaries, while the highest expression level in the testis is observed at 5 months (Figure 8D).

https://doi.org/10.1007/s11427-023-2694-5

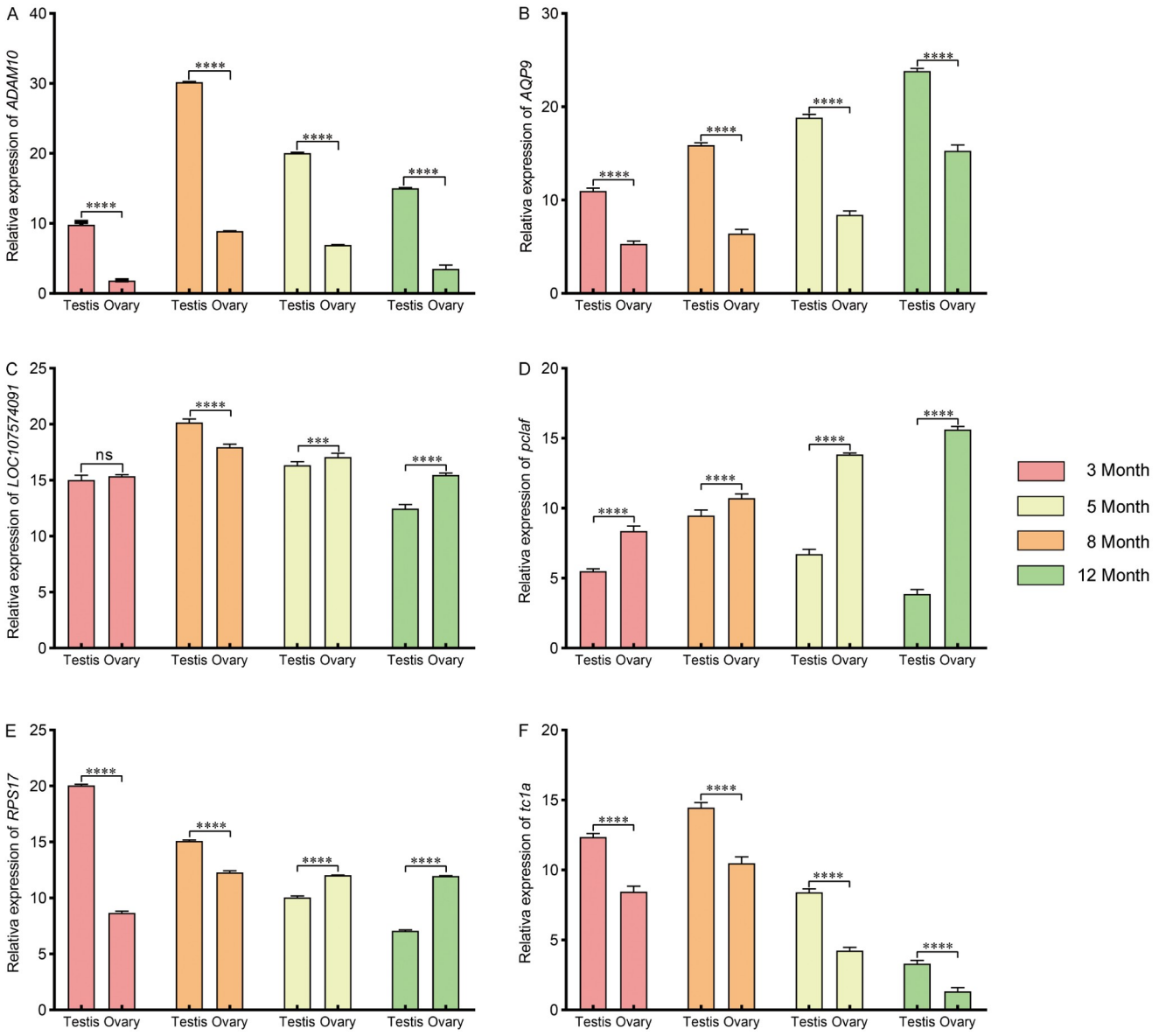

Figure 8. Verification of the expression of selected candidate sex determination genes *ADAM10* (A), *AQP9* (B), *LOC107574091* (C), *pclaf* (D), *RPS17* (E) and *tc1a* (F) by RT-qPCR. Error bars indicated the standard deviation of three biological replicates. The *y*-axis represents the expression level of these genes determined by RT-qPCR. * represents a significant difference with a *P*-value less than 0.05, ns represents no significant difference.

This phenomenon suggests that the *pclaf* gene may have different regulatory mechanisms in different gonadal tissues, but whether this is related to sex determination requires further research. The *tc1a* (Transposable element Tc1 transposase) gene is transposable element that belongs to the Tc1/mariner superfamily of transposon classes (Plasterk, 1996). Transposable elements (TEs) are repetitive sequences of DNA that can move within chromosomes or transpose between different sites in the genome (Ferreira et al., 2011). The insertion of TEs may lead to chromosomal rearrangements or mutations, thereby affecting the sex-determining function of sex chromosomes. Some studies have found that TEs, as part of chromosomal rearrangements, have significant effects on genome evolution (Harvey et al., 2002; Ozouf-Costaz et al., 2004; Steinemann and Steinemann, 2005). In this study, we found that the expression level of the gene in the testes was higher than in the ovaries, and it reached its highest level in 5-month-old testicular tissue (Figure 8F). The *tc1a* gene, which encodes the transposase for the transposable element Tc1, plays a crucial role in the transposition process of the Tc1 element. Previous studies have found that TEs had undergone bursts and simple repetitive accumulations in early sex determination regions (Chalopin et al., 2015). The gene exhibits its highest expression level in testicular tissue at 5 months of age, potentially associated with the significant increase in transposable elements (TEs), thereby further influencing the sex determination of 4nRR. In this study, we validated genes in the candidate sex-determining region across different developmental stages in gonadal tissues, providing a basis for further investigation into the sex determination of 4nRR. In the future, we still need to further explore the roles of these genes in the sex determination process of 4nRR using additional methods and approaches.

In summary, this study has yielded crucial insights and established methodologies for uncovering the sex determination mechanism of 4nRR (and potentially other related species) by developing male-specific molecular marker and identifying candidate sex determination genes through whole genome resequencing analysis. These findings significantly advance our understanding of the fundamental biology and provide critical tools for deciphering sex determination mechanisms in 4nRR and related species. The RT-qPCR validation further supports the correlation of these candidate genes with the sex determination process. These findings are significant not only in the field of fundamental biology but also hold potential applications in the aquaculture and genetic enhancement of 4nRR. Future research endeavors could delve deeper into elucidating the functional mechanisms of these candidate genes, along with exploring their expression dynamics across various developmental stages and environmental conditions. Moreover, in-depth investigation into these sex determination mechanisms is essential for advancing sex-controlled breeding strategies and fostering development in related fields.

MATERIAL AND METHODS

Ethics statement

The animal care and experimental protocols were certified by a professional training course for laboratory animal practitioners held by the Institute of Experimental Animals, Hunan Province, China. The experimental fish samples were deeply anesthetized with 100 mg L^{-1} MS-222 before collecting gonadal and muscle tissues.

Sample collection and DNA extraction

Ten females (FJF1–FJF50) and ten males (MJF1–MJF50) of 4nRR were collected from the State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, China. The phenotypic sex of individual was determined by anatomical examination of gonads. Muscle tissue collected was preserved using anhydrous ethanol and stored in –20 °C for genomic DNA extraction. Total genomic DNA was extracted from the muscle of 4nRR by using DNA extraction kit (Tsingke, China) according to the manufacturer's instructions. The NanoDrop 2000&8000 microvolume spectrophotometer (Thermo Fisher Scientific, USA) was used to measure the purity of the DNA. The Qubit 4.0 fluorometer (Thermo Fisher Scientific, USA) was used for DNA concentration measurement. The integrity test of DNA integrity was carried out through 1% agarose gel electrophoresis.

Genome re-sequencing and variants calling

The qualified DNA was utilized for constructing a genomic DNA library, followed by sequencing on the Illumina platform (NovaSeq 6000) using a paired-end 150 bp strategy according to standard manufacturer's protocols. The raw reads were quality controlled and filtered with fastp v0.23.4 (Chen et al., 2018). The clean reads were obtained after discarding the following: (1) the adapter sequence; (2) the reads with unknown nucleotides comprising more than 10%; (3) the low-quality reads being of more than 50% base with quality lower than 5.

The chromosome-level genomes of 4nRR (unpublished) was indexed and constructed using the index function of BWA v0.7.17 (Jung and Han, 2022). The clean reads were aligned to the reference genome of 4nRR by MEM algorithm of BWA and the alignment results were subsequently analyzed. The resulting alignments BAM files were sort and were converted into SAM files using the software samtools v1.5 (Danecek et al., 2021). The MarkDuplicates of Picard toolkit v3.0 (https://github.com/broadinstitute/picard) was used to remove duplicate sequences from the aligned clean reads and build index files of BAM files. The AddOrReplaceReadGroups of Picard toolkit was used to add or replace read groups information in SAM files. The HaplotypeCaller of GATK v4.0.4.0 (McKenna et al., 2010) was used to perform variant calling for each sample. VariantFiltration and SelectVariants of GATK v4.0.4.0 tools were used to extract SNPs and Indels from samples after variant detection. The filtering conditions for SNPs and Indels were as follows: QD<2.0||FS>200.0||MQ<40.0||MQRankSum<–12.5||ReadPosRankSum<–8.0. The number of SNPs and Indels with sexual dimorphism was counted per 1 Mb window, and visualized in a genome-wide density plot by CMplot package (Yin et al., 2021).

Genome-wide association studies and identification of sex-associated locus

VCF2PCACluster v1.38 (https://github.com/hewm2008/VCF2PCACluster) was used to perform the PCA (principal components analysis) based on filtered SNP. VCF2Dis was used to analyze the genetic relationship between male and female populations, and obtain a genetic relationship matrix between the two populations. And then, the ML (Maximum Likelihood) tree was constructed using IQ-TREE v2.2.5 (Nguyen et al., 2015), and then visualized with iTOL v6.8 (Letunic and Bork, 2021). The F_{st} value between the two populations was calculated using vcftools v4.0 (Danecek et al., 2011). Using the –freq of vcftools software, calculate the SNP frequencies between groups separately, quantify the differences in SNP frequencies between groups, and employ the ED (Euclidean distance) calculation commonly used in Bulk Segregant Analysis (BSA) to compute the Euclidean distances between groups. Counting the total number of SNPs between groups with vcftools, where the absolute frequency difference is greater than 0.5 and the number of SNPs exceeds 200. The above statistical results are visualized across the entire genome using the R package ggplot2 (Wickham, 2011). Based on the above three results, further determination of the sex chromosomes and sex-determining candidate regions in 4nRR were carried out.

Validation of male chromosome-specific fragments with PCR amplification

Visualizing SNPs and Indels in deduplicated BAM files using IGV (Integrative Genomic Viewer) v2.16.2 (Thorvaldsdóttir et al., 2013) to inspect differences between male and female populations. And then, one Indel marker was selected to develop sex-specific marker for 4nRR. The sex-specific marker was designed base on Indel sequence with Primer Premier v6.0 (Singh et al., 1998), and then synthesized by a service company (Tsingke, Beijing, China). Ten male and ten female fish were verified by PCR/gelelectrophoresis for the first time. For PCR amplification, a total reaction volume of 25 μL was prepared, containing 12.5 μL 2× Rapid Taq Master Mix (Vazyme Biotech, Nanjing; containing

https://doi.org/10.1007/s11427-023-2694-5

electrophoresis buffer and green loading buffer); 1 μL of forward and reverse primer mix (10 μmol L^{-1}); 1 μL of 20 ng μL^{-1} template genomic DNA and 9.5 μL of ddH_2O (double distilled water). A 96-well thermal cycler (SimpliAmp, Thermofisher) was used for PCR using the following program: pre-denaturation was performed at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 15 s, annealing at 46°C for 15 s, extension at 72°C for 15 s, and a final extension step at 72°C for 5 min. The PCR products are stored at 4°C for subsequent agarose gel electrophoresis detection. The 5 μL amplified products were separated by 2% agarose gel electrophoresis using the 100 bp DNA Ladder marker (Takara, Japan) as the standard DNA ladder. To further verify that this marker was truly sex-specific and not just specific to a few individuals, all 100 sampled fish of known sex were used to verify their detection efficiency.

Functional annotation of candidate genes for sex determination

Based on the selected sex candidate regions determined from the gene annotation information in the gff file, a total of six genes that meet the criteria were identified. To further investigate the linkage disequilibrium (LD) between loci within the sex-determination region, LDBlockShow v1.4 (Dong et al., 2021) software was employed to assess the non-random association among different loci in the genome where these six candidate genes are located. Next, extract the coding protein sequences of these six genes from the genomic CDS (Coding DNA Sequence) sequence file. The six candidate genes were aligned to nonredundant protein sequences (NRs) using Diamond v0.8.22.84 (Buchfink et al., 2021) with a cutoff *e*-value of 1×10^{-5}. Extracted six protein sequences and used the eggNOG-mapperb (Cantalapiedra et al., 2021) and UniProt online annotation tool to convert them into Gene symbols. Utilize the online gprofiler (Kolberg et al., 2023) tool for converting gene symbols to Entrez Gene IDs, followed by using KOBAS (Bu et al., 2021) for online gene functional annotation and functional enrichment analysis. The R packages clusterProfiler (Yu et al., 2012) was used for GO (Gene ontology) and KEGG (Kyoto encyclopedia of genes and genomes) functional enrichment analysis.

Validation of candidate genes by RT-qPCR

Three males and three females (3, 5, 8, and 12 month) were used for RNA extraction to validate candidate genes. Total RNA was extracted from ovary and testis samples using TaKaRa MiniBEST Universal RNA Extraction Kit (Takara, Japan) according to the manufacturer's instructions. Extracted RNA was reverse transcribed into cDNA using the Maxima H Minus First Strand cDNA Synthesis Kit with dsDNase (Thermo Scientific, China) reagent according to the manufacturer's instructions. Utilizing the chromosomal genomic coordinates of the gene, we employed the faidx function in samtools v1.5 (Danecek et al., 2021) to extract the gene sequence. To ensure primer specificity, we conducted a comprehensive genome alignment using BLAST v2.7.1+ (Ye et al., 2006) to identify repetitive sequences associated with the gene. The RT-qPCR primers for these genes were designed using Primer Premier v6.0 (Singh et al., 1998) software and were listed in Table S6. The reaction system for RT-qPCR is as follows: 2 μL of cDNA; 0.4 μL each of forward and reverse primers; 2.2 μL of ddH_2O; 5 μL of 2X ChamQ Universal SYBR PCR Master Mix (Takara, Japan). The *β-actin* gene was used as an internal control to normalize transcript levels. Reactions were performed with the following conditions: 95 °C for 30 s, 40 cycles of 95 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s. All samples were analyzed in triplicate and the expression of target genes was calculated as relative folds with the $2^{-\triangle\triangle C_T}$ method. All data were expressed as mean±standard deviation. Student *t*-test was used to analyze the statistical differences at 0.05.

Availability of data and materials

The raw genome sequencing data used in this study have been were deposited in China National Center Bioinformation (CNCB) Genome Sequence Archive (https://www.cncb.ac.cn/. Accession number: CRX805054, CRX805053, CRX805052, CRX805051, CRX805050, CRX805049, CRX805048, CRX805047, CRX805046, CRX805045, CRX805044, CRX805043, CRX805042, CRX805041, CRX805040, CRX805039, CRX805038, CRX805037, CRX805036, and CRX805035). The project number for the remaining eighty raw data entries is PRJCA027165.

Compliance and ethics

The authors declare that they have no competing interests.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (32172972), the Science and Technology Innovation Program of Hunan Province (2021RC4028), and the Special Funds for Construction of Innovative Provinces in Hunan Province (2021NK1010), Special Science Found of Nansha-South China Agricultural University Fishery Research Institute, Guangzhou, the Aid Program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province, the earmarked fund for HARS (HARS-07). We thank Wuhan Onemore-tech Co., Ltd. for their assistance with genome sequencing and analysis.

Supporting information

The supporting information is available online at https://doi.org/10.1007/s11427-023-2694-5. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: why so many ways of doing it? PLoS Biol 12, e1001899.

Blobel, C.P. (1997). Metalloprotease-disintegrins: links to cell adhesion and cleavage of TNFα and notch. Cell 90, 589–592.

Blobel, C.P. (2005). ADAMs: key components in EGFR signalling and development. Nat Rev Mol Cell Biol 6, 32–43.

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., et al. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res 49, W317–W325.

Buchfink, B., Reuter, K., and Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18, 366–368.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., and Tamura, K. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol 38, 5825–5829.

Chalopin, D., Volff, J.N., Galiana, D., Anderson, J.L., and Schartl, M. (2015). Transposable elements and early evolution of sex chromosomes in fish. Chromosome Res 23, 545–560.

Chen, J., Shao, X., Ye, S., Xiao, J., Zou, Y., Ye, K., Zhang, L., Wang, Z., Xiao, S., and Cai, M. (2022). Development of sex-specific markers in Spinyhead croaker, *Collichthys lucidus*. Aquaculture 547, 737424.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, S.L., Ji, X.S., Shao, C.W., Li, W.L., Yang, J.F., Liang, Z., Liao, X.L., Xu, G.B., Xu, Y., and Song, W.T. (2012). Induction of mitogynogenetic diploids and identification of WW super-female using sex-specific SSR markers in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol 14, 120–128.
Chen, S.L., Deng, S.P., Ma, H.Y., Tian, Y.S., Xu, J.Y., Yang, J.F., Wang, Q.Y., Ji, X.S., Shao, C.W., Wang, X.L., et al. (2008). Molecular marker-assisted sex control in half-smooth tongue sole (*Cynoglossus semilaevis*). Aquaculture 283, 7–12.
Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience 10, 1–4.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158.
Devlin, R.H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. Aquaculture 208, 191–364.
Devlin, R.H., McNeil, B.K., Groves, T.D.D., and Donaldson, E.M. (1991). Isolation of a Y-Chromosomal DNA Probe Capable of Determining Genetic Sex in Chinook Salmon (*Oncorhynchus tshawytscha*). Can J Fish Aquat Sci 48, 1606–1612.
Dong, S.S., He, W.M., Ji, J.J., Zhang, C., Guo, Y., and Yang, T.L. (2021). LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Brief BioInf 22, bbaa227.
Du, J., Zhou, J., Li, S., Shao, J., Jiang, P., Dong, C., and Bai, J. (2021). A PCR-based method for genetic sex identification and evidence of the XX/XY sex determination system in largemouth bass (*Micropterus salmoides* L.). Aquaculture 545, 737220.
Ellegren, H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet 12, 157–166.
Ezaz, M.T., Harvey, S.C., Boonphakdee, C., Teale, A.J., McAndrew, B.J., and Penman, D.J. (2004). Isolation and physical mapping of sex-linked AFLP markers in Nile Tilapia (*Oreochromis niloticus* L.). Mar Biotechnol 6, 435–445.
Ferreira, D.C., Porto-Foresti, F., Oliveira, C., and Foresti, F. (2011). Transposable elements as a potential source for understanding the fish genome. Mobile Genet Elem 1, 112–117.
Gammerdinger, W.J., Conte, M.A., Baroiller, J.F., D'Cotta, H., and Kocher, T.D. (2016). Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. BMC Genomics 17, 808.
Gao, F.X., Shi, Y., Duan, W., Lu, W.J., Huang, W., Zhang, X.J., Zhao, Z., Zhou, L., and Gui, J.F. (2020). A rapid and reliable method for identifying genetic sex in obscure pufferfish (*Takifugu obscurus*). Aquaculture 519, 734749.
Griffiths, R., Orr, K.L., Adam, A., and Barber, I. (2000). DNA sex identification in the three-spined stickleback. J Fish Biol 57, 1331–1334.
Han, C., Zhou, X., Lu, H., Zhu, Q., Han, L., Li, S., Lin, H., and Zhang, Y. (2021). A simple PCR-based genetic sex identification method in the blotched snakehead (*Channa maculata*) developed by high-throughput sequencing. Aquaculture 538, 736579.
Harvey, S.C., Campos-Ramos, R., Kennedy, D.D., Ezaz, M.T., Bromage, N.R., Griffin, D. K., and Penman, D.J. (2002). Karyotype evolution in tilapia: mitotic and meiotic chromosome analysis of *Oreochromis karongae* and *O. Niloticus* × *O. Karongae* hybrids. Genetica 115, 169–177.
Huang, H.F., He, R.H., Sun, C.C., Zhang, Y., Meng, Q.X., and Ma, Y.Y. (2006). Function of aquaporins in female and male reproductive systems. Hum Reprod Update 12, 785–795.
Huovila, A.P.J., Almeida, E.A., and White, J.M. (1996). ADAMs and cell fusion. Curr Opin Cell Biol 8, 692–699.
Jung, Y., and Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics 38, 2404–2413.
Kageyama, Y., Ishibashi, K., Hayashi, T., Xia, G., Sasaki, S., and Kihara, K. (2001). Expression of aquaporins 7 and 8 in the developing rat testis. Andrologia 33, 165–169.
Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). PLoS Genet 8, e1002798.
Kobayashi, Y., Nagahama, Y., and Nakamura, M. (2012). Diversity and plasticity of sex determination and differentiation in fishes. Sex Dev 7, 115–125.
Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g: Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res 51, W207–W212.
Kovács, B., Egedi, S., Bártfai, R., and Orbán, L. (2000). Genetica 110, 267–276.
Kuang, G., Tao, W., Zheng, S., Wang, X., and Wang, D. (2020). Genome-wide identification, evolution and expression of the complete set of cytoplasmic ribosomal protein genes in Nile Tilapia. Int J Mol Sci 21, 1230.
Lee, B.Y., Coutanceau, J.P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J.F., and Kocher, T. D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile Tilapia (*Oreochromis niloticus*). Mar Biotechnol 13, 557–562.
Leinonen, T., Cano, J.M., and Merilä, J. (2011). Genetic basis of sexual dimorphism in the threespine stickleback *Gasterosteus aculeatus*. Heredity 106, 218–227.
Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49, W293–W296.
Li, Q., Xie, J., He, L., Wang, Y., Duan, Z., Yang, H., and Wang, Q. (2015). Identification of ADAM10 and ADAM17 with potential roles in the spermatogenesis of the Chinese mitten crab, *Eriocheir sinensis*. Gene 562, 117–127.
Li, X.Y., and Gui, J.F. (2018). Diverse and variable sex determination mechanisms in vertebrates. Sci China Life Sci 61, 1503–1514.
Li, X.Y., Mei, J., Ge, C.T., Liu, X.L., and Gui, J.F. (2022). Sex determination mechanisms and sex control approaches in aquaculture animals. Sci China Life Sci 65, 1091–1122.
Liang, Y., Meyer, A., and Kratochwil, C.F. (2020). Neural innervation as a potential trigger of morphological color change and sexual dimorphism in cichlid fish. Sci Rep 10, 12329.
Lin, A., Xiao, S., Xu, S., Ye, K., Lin, X., Sun, S., and Wang, Z. (2017). Identification of a male-specific DNA marker in the large yellow croaker (*Larimichthys crocea*). Aquaculture 480, 116–122.
Lin, X.Y., Xiao, S.J., Li, W.B., and Wang, Z.Y. (2018). Development and validation of sex-specific markers in *Piaractus mesopotamicus*. J Fisher China 42, 1329–1337.
Liu, S., Han, C., Huang, J., Zhu, Q., Liu, D., Han, L., Li, S., Li, G., Lin, H., and Zhang, Y. (2023). Screening and characterization of X chromosome-specific markers in mandarin fish (*Siniperca chuatsi*). Aquaculture 562, 738833.
Mascali, F.C., Posner, V.M., Romero Marano, E.A., del Pazo, F., Hermida, M., Sánchez, S., Mazzoni, T.S., Martinez, P., Rubiolo, J.A., and Villanova, G.V. (2022). Development and validation of sex-specific markers in *Piaractus mesopotamicus*. Aquaculture 558, 738374.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303.
Mei, J., and Gui, J.F. (2015). Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. Sci China Life Sci 58, 124–136.
Monsang, S.J., Parhi, J., Choudhury, J., Deka, A., and Pal, P. (2019). Molecular characterisation of carp aquaporin-3a: expression profiling during hormone-induced spawning. Aquaculture 500, 569–575.
Nagahama, Y., Chakraborty, T., Paul-Prasanth, B., Ohta, K., and Nakamura, M. (2020). Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. Physiol Rev 101, 1237–1308.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32, 268–274.
Otto, S.P. (2007). The evolutionary consequences of polyploidy. Cell 131, 452–462.
Ozouf-costaz, C., Brandt, J., Körting, C., Pisano, E., Bonillo, C., Coutanceau, J.P., and Volff, J.N. (2004). Genome dynamics and chromosomal localization of the non-LTR retrotransposons *Rex1* and *Rex3* in Antarctic fish. Antartic Sci 16, 51–57.
Peng, Y.X., Huang, Y.Q., Zhong, J., Jiang, Z.T., Fan, S., Shi, H.J., Chen, H.P., Deng, S. P., Li, G.L., and Jiang, D.N. (2023). Identification of sex-linked marker and candidate sex determination gene in ornamental fish, African scat (*Scatophagus tetracanthus*). Aquaculture 563, 739023.
Perez-Enriquez, R., Llera-Herrera, R., Galindo Torres, P., and Avila, S. (2020). A fast sex detection method for the Whiteleg shrimp *Litopenaeus vannamei* by post-PCR high resolution melting (HRM). Aquaculture 520, 734784.
Plasterk, R.H.A. (1996). The Tc1/mariner transposon family. Curr Top Microbiol Immunol 204, 125–143.
Qin, Q., Wang, Y., Wang, J., Dai, J., Xiao, J., Hu, F., Luo, K., Tao, M., Zhang, C., Liu, Y., et al. (2014). The autotetraploid fish derived from hybridization of *Carassius auratus* red var. (Female) × Megalobrama amblycephala (Male)1. Biol Reprod 91, 93.
Reddon, A.R., and Hurd, P.L. (2013). Water pH during early development influences sex ratio and male morph in a West African cichlid fish, *Pelvicachromis pulcher*. Zoology 116, 139–143.
Ruan, R., Feng, T., Li, Y., Yue, H., Ye, H., Du, H., Liu, Q., Ruan, J., Li, C., and Wei, Q. (2021). Screening and identification of female-specific DNA sequences in octaploid sturgeon using comparative genomics with high-throughput sequencing. Genomics 113, 4237–4244.
Sandra, G.E., and Norma, M.M. (2010). Sexual determination and differentiation in teleost fish. Rev Fish Biol Fisheries 20, 101–121.
Seals, D.F., and Courtneidge, S.A. (2003). The ADAMs family of metalloproteases: multidomain proteins with multiple functions. Genes Dev 17, 7–30.
Shen, Z.G., and Wang, H.P. (2014). Molecular players involved in temperature-dependent sex determination and sex differentiation in Teleost fish. Genet Sel Evol 46, 26.
Shi, X., Waiho, K., Li, X., Ikhwanuddin, M., Miao, G., Lin, F., Zhang, Y., Li, S., Zheng,

H., Liu, W., et al. (2018). Female-specific SNP markers provide insights into a WZ/ZZ sex determination system for mud crabs *Scylla paramamosain*, *S. tranquebarica* and *S. serrata* with a rapid method for genetic sex identification. BMC Genomics 19, 981.
Shikano, T., Herczeg, G., and Merilä, J. (2011). Molecular sexing and population genetic inference using a sex-linked microsatellite marker in the nine-spined stickleback (*Pungitius pungitius*). BMC Res Notes 4, 119.
Simpson, F., Lammertsvanbueren, K., Butterfield, N., Bennetts, J., Bowles, J., Adolphe, C., Simms, L., Young, J., Walsh, M., and Leggett, B. (2006). The PCNA-associated factor KIAA0101/p15 binds the potential tumor suppressor product p33ING1b. Exp Cell Res 312, 73–85.
Singh, V.K., Mangalam, A.K., Dwivedi, S., and Naik, S. (1998). Primer premier: program for design of degenerate primers from a protein sequence. BioTechniques 24, 318–319.
Small, C.M., Carney, G.E., Mo, Q., Vannucci, M., and Jones, A.G. (2009). A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: evidence for masculinization of the transcriptome. BMC Genomics 10, 579.
Steinemann, S., and Steinemann, M. (2005). Retroelements: tools for sex chromosome evolution. Cytogenet Genome Res 110, 134–143.
Sun, Y.L., Jiang, D.N., Zeng, S., Hu, C.J., Ye, K., Yang, C., Yang, S.J., Li, M.H., and Wang, D.S. (2014). Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Nile tilapia (*Oreochromis niloticus*). Aquaculture 433, 19–27.
Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief BioInf 14, 178–192.
Vale, L., Dieguez, R., Sánchez, L., Martínez, P., and Viñas, A. (2014). A sex-associated sequence identified by RAPD screening in gynogenetic individuals of turbot (*Scophthalmus maximus*). Mol Biol Rep 41, 1501–1509.
Wang, C., Qin, H., Zhao, C., Yang, L., Yu, T., Zhang, Y., Luo, X., Qin, Q., and Liu, S. (2021). Whole-genome re-sequencing and transcriptome reveal oogenesis-related genes in autotetraploid *Carassius auratus*. Mar Biotechnol 23, 233–241.
Wang, D., Mao, H.-., Chen, H.-., Liu, H.-., and Gui, J.-. (2009). Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. anim Genet 40, 978–981.
Weng, X., Xu, Y., Dong, X., Luo, X., You, W., Ke, C., and Cai, M. (2022). Sex-specific markers developed by next-generation sequencing confirmed a male heterogametic sex determination in small abalone, *Haliotis diversicolor*. Aquaculture 555, 738256.
Wickham, H. (2011). ggplot2. WIREs Comput Stats 3, 180–185.
Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci USA 106, 13875–13879.
Xiao, J., Zou, Y., Xiao, S., Chen, J., Wang, Z., Wang, Y., Jie, X., and Cai, M. (2020). Development of a PCR-based genetic sex identification method in spinyhead croaker (*Collichthys lucidus*). Aquaculture 522, 735130.
Yamamoto, Y., Hattori, R.S., Patiño, R., Strüssmann, C.A. Chapter Two - Environmental regulation of sex determination in fishes: insights from Atheriniformes. In: Capel B, ed. Current Topics in Developmental Biology. New York: Academic Press, 2019.
Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Curr Biol 22, 1423–1428.
Ye, J., McGinnis, S., and Madden, T.L. (2006). BLAST: Improvements for better sequence analysis. Nucleic Acids Res 34, W6–W9.
Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., et al. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics BioInf 19, 619–628.
Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS-J Integrative Biol 16, 284–287.
Zhang, A., Huang, R., Chen, L., Xiong, L., He, L., Li, Y., Liao, L., Zhu, Z., and Wang, Y. (2017). Computational identification of Y-linked markers and genes in the grass carp genome by using a pool-and-sequence method. Sci Rep 7, 8213.
Zhang, B., Zhao, N., Liu, Y., Jia, L., Fu, Y., He, X., Liu, K., Xu, Z., and Bao, B. (2019). Novel molecular markers for high-throughput sex characterization of *Cynoglossus semilaevis*. Aquaculture 513, 734331.
Zhang, K., Song, F., Liang, Y., Zhang, W., Zhou, K., Chen, Y., Zhen, P., and Luo, J. (2022). Development and verification of sex-specific molecular marker for Golden pompano (*Trachinotus blochii*). Aquaculture Res 53, 3726–3735.
Zheng, S., Wang, X., Zhang, S., Long, J., Tao, W., Li, M., and Wang, D. (2020). Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Southern catfish (*Silurus meridionalis*). Aquaculture 517, 734783.
Zhou, Y., Wu, J., Wang, Z., Li, G., Mei, J., Zhou, L., and Gui, J. (2019). Identification of sex-specific markers and heterogametic XX/XY sex determination system by 2b-RAD sequencing in redtail catfish (*Mystus wyckioides*). Aquac Res 50, 2251–2266.
Zhu, C., Liu, H., Cheng, L., Pan, Z., Chang, G., Wu, N., Wang, H., Ding, H., Fang, Y., and Wang, L. (2021). Identification of sex-specific sequences through 2b-RAD sequencing in *Pseudobagrus ussuriensis*. Aquaculture 539, 736639.