PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Zheng S *et al*. 2022 Identification of sex chromosome and sex-determining gene of southern catfish (*Silurus meridionalis*) based on XX, XY and YY genome sequencing. *Proc. R. Soc. B* **289**: 20212645.
https://doi.org/10.1098/rspb.2021.2645

Received: 8 December 2021
Accepted: 15 February 2022

Subject Category:
Genetics and genomics

Subject Areas:
genetics, genomics, evolution

Keywords:
southern catfish, Nanopore sequencing, Hi-C, Y chromosome, sex determination gene, *amhr2y*

Authors for correspondence:
Yaoguang Zhang
e-mail: zhangyg@swu.edu.cn
Deshou Wang
e-mail: wdeshou@swu.edu.cn

†Shuqing Zheng and Wenjing Tao contributed equally to this study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5877641.

THE ROYAL SOCIETY PUBLISHING

Identification of sex chromosome and sex-determining gene of southern catfish (*Silurus meridionalis*) based on XX, XY and YY genome sequencing

Shuqing Zheng[1,†], Wenjing Tao[1,†], Haowen Yang[1], Thomas D. Kocher[2], Zhijian Wang[1], Zuogang Peng[1], Li Jin[1], Deyong Pu[1], Yaoguang Zhang[1] and Deshou Wang[1]

[1]Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, 400715 Chongqing, People's Republic of China
[2]Department of Biology, University of Maryland, College Park, MD 20742, USA

SZ, 0000-0003-1007-6116; TDK, 0000-0002-7547-0133; ZW, 0000-0001-9826-2399; ZP, 0000-0001-8810-2025; DP, 0000-0002-7971-0130; YZ, 0000-0003-1695-1890; DW, 0000-0002-4967-524X

Teleosts are important models to study sex chromosomes and sex-determining (SD) genes because they present a variety of sex determination systems. Here, we used Nanopore and Hi-C technologies to generate a high-contiguity chromosome-level genome assembly of a YY southern catfish (*Silurus meridionalis*). The assembly is 750.0 Mb long, with contig N50 of 15.96 Mb and scaffold N50 of 27.22 Mb. We also sequenced and assembled an XY male genome with a size of 727.2 Mb and contig N50 of 13.69 Mb. We identified a candidate SD gene through comparisons to our previous assembly of an XX individual. By resequencing male and female pools, we characterized a 2.38 Mb sex-determining region (SDR) on Chr24. Analysis of read coverage and comparison of the X and Y chromosome sequences showed a Y specific insertion (approx. 500 kb) in the SDR which contained a male-specific duplicate of *amhr2* (named *amhr2y*). *amhr2y* and *amhr2* shared high-nucleotide identity (81.0%) in the coding region but extremely low identity in the promotor and intron regions. The exclusive expression in the male gonadal primordium and loss-of-function inducing male to female sex reversal confirmed the role of *amhr2y* in male sex determination. Our study provides a new example of *amhr2* as the SD gene in fish and sheds light on the convergent evolution of the duplication of AMH/AMHR2 pathway members underlying the evolution of sex determination in different fish lineages.

1. Background

Sex chromosomes carry a sex-determining (SD) gene that initiates gonadal differentiation in the earliest stages of development. Teleosts are important models to study sex chromosomes and SD genes because they present a variety of sex-determination systems. There are two main categories for the production of SD genes, namely allelic diversification and gene duplication [1]. To date, teleosts are the only group where both categories have been found [2–4]. Most of the identified SD genes in teleosts belong to either the SOX or DMRT gene families, or a member of the TGF-β signalling pathway (reviewed in [5]). This suggests that only a limited group of factors/signalling pathways are prone to become top regulators. The TGF-β signalling pathway is particularly noteworthy since several members of this pathway (*amh*, *amhr2*, *gsdf*, *gdf6*, *bmpr1b*) have been recruited as SD genes in about 16 fish species. However, due to the relatively small number of species with known/isolated SD genes, the role of TGF-β signalling pathway members as SD genes in fish is likely underestimated.

Highly differentiated Y or W chromosomes with conserved SD genes have been found in mammals and birds [6,7]. By stark contrast, most teleosts have homomorphic sex chromosomes and display a high diversity of SD genes [8]. Only about 10% of the teleosts that have been characterized possess heteromorphic sex chromosomes [9]. Although SD gene can be mapped to a specific chromosomal region by genetic analysis, identification of the SD gene in most teleosts is still a daunting task. At present, the SD gene has been identified in fewer than 25 species, just the tip of the iceberg in a sea of 20 000 teleosts [5].

Southern catfish (*Silurus meridionalis*) is an economically important silurid fish which is widely distributed in the Yangtze River basin of China. On the one hand, it has emerged as an important fish species for ecology and conservation because of its pivotal role as a top predator that shapes the structure of local fish communities. On the other hand, it is an excellent model for studying sexual dimorphism and sex determination in fish as adult females grow as much as two-fold bigger than the males [10]. Southern catfish has $2n = 58$ chromosomes and the sex chromosome pair is homomorphic [11]. Our previous study identified sex-linked DNA markers and demonstrated an XX/XY sex-determination system in this species [12]. We also generated a whole-genome assembly anchored on chromosomes for an XX female individual [13]. Yet, little is known about genetic sex determination in southern catfish beyond males being the heterogametic sex, and its sex locus and SD gene remain elusive.

As the SD gene is always located on the Y or W chromosome no matter how it is produced, generation and sequencing of a YY/WW individual is undoubtedly a fast and effective way to isolate SD gene. In this study, we sequenced and assembled the genomes of an XY male and a YY supermale by Nanopore and Hi-C. We then used the whole-genome sequencing data of XX, XY and YY, together with the pool resequencing data from 110 males and 41 females, to characterize a Y-specific insertion on Chr24 harbouring the candidate SD gene *amhr2y*. The role of this gene in male sex determination was demonstrated by substantial expression analysis and by functional analysis of CRISPR knockouts.

2. Material and methods

(a) DNA sampling and genome sequencing

Adult southern catfish were obtained from the Jialing River and were reared at the Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Chongqing, China. Progeny were obtained by artificial fertilization. Based on our previous study, feeding *Limnodilus* spp., an annelid found abundantly in organically polluted waters, from 1 to 30 dah (days after hatching) resulted in complete feminization of fry in southern catfish, which has a 1 : 1 sex ratio in the wild [14]. In this study, XY neofemales were crossed to normal XY males, which produced offspring with a sex ratio of 3♂ : 1♀ (YY: XY: XX = 1 : 2 : 1) when fed commercial diets (Shengsuo, China). An XY male and a YY supermale identified by sex-linked markers [12] were used for DNA sequencing.

Two Illumina libraries with insert sizes of 350 bp were constructed and then sequenced on an MGISEQ-2000 instrument. Raw reads were filtered using the Trimmomatic v. 0.33 (parameters: PE -threads 4 -phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 MINLEN:15) to remove adaptors and low-quality reads [15]. Two Nanopore libraries were constructed and sequenced on two different FlowCells using the PromethION sequencer (ONT, UK). To anchor the contigs onto the chromosome, genomic DNA was extracted from the blood of YY southern catfish for the Hi-C library as described previously [16], and sequencing on HiSeq X Ten platform (Illumina, USA) was performed using a PE-150 module.

(b) Pool resequencing

A female pool (F-pool, including 41 female individuals) and a male pool (M-pool, including 110 male individuals) were used for DNA resequencing. The phenotypic sex was identified by gonad histological examination. Genomic DNA extracted from fin clips was pooled in equimolar ratio according to sex and pool-sequencing libraries were prepared according to Illumina's protocols. After quality inspection and concentration measurement, the library was paired-end sequenced (2×150 bp) on Illumina HiSeq X platform (Illumina, USA).

(c) Genome assembly

The Nanopore sequencing data were corrected and assembled using Nextdenovo v. 2.5.0 (https://github.com/Nextomics/NextDenovo) and polished using Nanopolish v. 0.13.0 and Pilon v. 1.22 [17,18]. For Hi-C assembly, the contigs were clustered, ordered and oriented onto chromosomes using Juicer and 3D-DNA pipeline with default parameters [19,20]. Genome completeness was estimated using BUSCO v. 3.0.1 [21] based on 4584 orthologues derived from the Actinopterygii lineage.

(d) Identification of sex-determining region by single nucleotide polymorphism (SNP) calling

The clean reads from F-pool and M-pool were aligned to the southern catfish YY genome using the BWA *mem* [22]. The BAM files were sorted using Samtools v. 1.3.1 [23] and PCR duplicates removed using Picard v. 2.18.2 (http://broadinstitute.github.io/picard/). Variants were called with the GATK pipeline [24]. Sex_SNP_finder_GA.pl (https://github.com/Gammerdinger/sex-SNP-finder) was used to identify sex-patterned SNPs and to calculate F_{ST} values. The sex-determining region (SDR) was defined by identifying non-overlapping 10 kb windows containing at least 50 sex-patterned SNPs. The non-overlapping window did not include positions with coverage less than 15 reads in both sexes.

(e) Identification of Y-specific region and candidate sex-determining gene

XX, XY and YY ONT long reads were mapped to the YY genome using Minimap2 v. 2.17-r941 [25] with default parameters, and the Illumina reads from F-pool and M-pool were mapped to the YY genome using BWA mem v. 0.7.12 [22]. SAMtools and Deeptools were used to calculate the coverage information [26]. IGV v2.5.2 [27] was used to show the coverage differences of reads from different sexes on the reference genome. Homology-based prediction was conducted using GeMoMa v. 1.6.2 [28]. Genome fasta and gff files of channel catfish (*Ictalurus punctatus*, GCA_001660625.1) was downloaded from NCBI. To validate the sex-linkage of *amhr2y* in males, five pairs of *amhr2y* specific primers and one pair of common primer were designed based on the alignment of *amhr2* and *amhr2y* (electronic supplementary material, table S1). The phenotypic sex was identified by histological examination. Genotyping was carried out on 75 gDNA samples using a PCR approach.

(f) Expression analyses

Total RNA was extracted from 12 tissues of pooled three XX and three XY fish at 120 dah. After DNase I (RNase free) treatment,

total RNA (1 µg) from each sample was reverse transcribed into first-strand cDNA and RT-PCR was performed to reveal the tissue distribution expression patterns of *amhr2y* and *amhr2*. Positive and negative controls were set up with plasmid DNA and negative control cDNA, respectively. *eef1a1a* was used as an internal control. Primer sequences used for RT-PCR are listed in the electronic supplementary material, table S1.

Three pairs of RNA preparations from gonads of XX and XY southern catfish at 120, 360 and 540 dah were sequenced using Illumina HiSeq2000 platform. Clean reads from each library were aligned to the reference genome using STAR v2.7.3a [29]. The FPKM (fragments per kilobase of exon per million reads mapped) method was used to calculate the gene expression level by RSEM [30].

The gonadal morphology between male and female was distinguishable at 10 dah. The gonadal transverse section of female was tubby and the male was spindly. In addition, the formation of the ovarian cavity (OC) at 12 dah could be a reliable criterion for identifying the gonad as an ovary [14]. To analyse the expression and cellular localization of *amhr2* and *amhr2y* in early developmental stages of male and female gonads, fluorescent *in situ* hybridization (FISH) was performed using southern catfish gonads at 5, 10, 30 and 120 dah. Alignment of *amhr2y* and *amhr2* revealed the highest sequence divergence at the 5′-end (approx. 25%) of the two cDNAs (electronic supplementary material, figure S2). Therefore, the 5′-end approximately 500 bp sequences of the two cDNAs were amplified with gene-specific primers to prepare probes (electronic supplementary material, table S1). Probes of sense and antisense digoxigenin labelled RNA strands were transcribed *in vitro* using a RNA labelling kit (Roche, Germany). Primers used in this section are listed in the electronic supplementary material, table S1. The fixation, embedding, sectioning of dissected gonads and FISH were performed as described previously [31].

(g) Disruption of *amhr2y* by CRISPR/Cas9

CRISPR/Cas9 was performed to mutation *amhr2y* in Southern catfish as described previously [32]. The gRNA of *amhr2y* and Cas9 mRNA were co-injected into one-cell stage embryos at a final concentration of 250 and 500 ng µl^{-1}, respectively. In order to assess the mutations, gDNA was extracted from pooled control and injected embryos (20 embryos for each) at 72 h after injection. DNA fragments spanning the target site were amplified. The mutated sequences were analysed by *Tsp*RI digestion and Sanger sequencing. Considering the high-sequence homology between *amhr*2y and *amhr2*, the off-target effect was detected by PCR and Sanger sequencing with *amhr2*-specific primers (electronic supplementary material, table S1).

At 90 dah, mutated fishes were identified by restriction enzyme digestion and Sanger sequencing as described above. The DNA samples extracted from tail fins of each fish were used for mutation analysis. The percentage of uncleaved band was measured by quantifying the band intensity of the *Tsp*RI digestion with Quantity One Software (Bio-Rad, USA). In addition, the percentage of the frame shifts was assessed by counting the number of frame shift mutations in 30 transformed recombinant clones per individual fish. Gonads to be processed for histology were fixed immediately after dissection in Bouin's fixative solution for 24 h. Samples were dehydrated and embedded in paraffin and then were cut serially into slices of 5 µm using a Leica microtome (Leica Microsystems, Germany) and stained with HE. Images were taken under Olympus BX51 light microscope (Olympus, Japan).

(h) Phylogenetic and syntenic analyses

Phylogenetic tree of Amhr2/Amhr2y was constructed using the maximum-likelihood method and Jones–Taylor–Thornton model implemented in MEGA 6.0 [33]. Amhr2 protein sequences of 15 teleost species and spotted gar, which was used as an outgroup, were retrieved from the NCBI protein database (electronic supplementary material, table S2). For syntenic analysis, position and orientation of *amhr2*/*amhr2y* and their adjacent genes on chromosomes of different species were determined using NCBI and Genomicus [34].

3. Results

(a) Sequencing and assembly of the XY and YY genomes

From Nanopore sequencing, a total of 69.1 Gb and 77.1 Gb clean data were collected after quality filtering for XY and YY, respectively (electronic supplementary material, tables S3–S4). After Nextdenovo assembly, Nanopolish calibration and three rounds of Pilon assembly using 42.8 Gb XY and 39.2 Gb YY Illumina clean reads, the XY and YY genome assemblies were approximately 727.2 Mb and 750.0 Mb with a contig N50 of 13.81 Mb and 15.96 Mb, respectively (electronic supplementary material, tables S5 and S6). The completeness of the two assemblies was further evaluated by BUSCO. Approximately, 90.7% and 92.3% of the complete BUSCO (Actinopterygii set) were found in the XY and YY assemblies, respectively (electronic supplementary material, table S7).

Hi-C data of the YY individual were further used for chromosome assembly of YY genome. As a result, 739.1 Mb (covering 98.54% of contigs on the base level) was anchored and orientated onto 29 chromosomes with a scaffold N50 of 27.22 Mb (electronic supplementary material, table S8–S13), which was consistent with the chromosome numbers from karyotype analysis [11]. The heatmap of chromosome cross-talk illustrated the completeness and robustness of the genome assembly (figure 1*a*). Additionally, the assembled 29 chromosomes of YY genome and previously reported XX female genome [13] showed a full correspondence in the synteny analysis (figure 1*b*).

(b) Characterization of the sex chromosome and sex-determining region

Previously, we identified eight sex-linked DNA markers in southern catfish through bioinformatics analysis and PCR verification [12]. In order to locate their positions on the chromosomes, we aligned the sequences generated by these markers to the XX and YY genomes, respectively. All sex-linked sequences are located on Chr24, which indicated this chromosome is the southern catfish sex chromosome (electronic supplementary material, table S14). We then conducted a pool-seq analysis to characterize the SDR. After filtering and quality control, 51.2 Gb and 38.7 Gb of clean data were obtained for F-pool and M-pool, respectively (electronic supplementary material, table S15). The clean data of the two pools were mapped to the YY genome with the mapping rates of 95.90% and 98.96%, and the coverage depths of 68.29 × and 51.58 ×, respectively (electronic supplementary material, table S16). Based on the alignments, 2 421 301 and 2 468 613 SNPs were obtained in F-pool and M-pool, respectively (electronic supplementary material, table S17). When applying stringent criteria demanding that gender-specific SNPs should be heterozygous in M-pool and homozygous in F-pool, we identified a genomic region of approximately 2.38 Mb (from 3.75–6.13 Mb) on

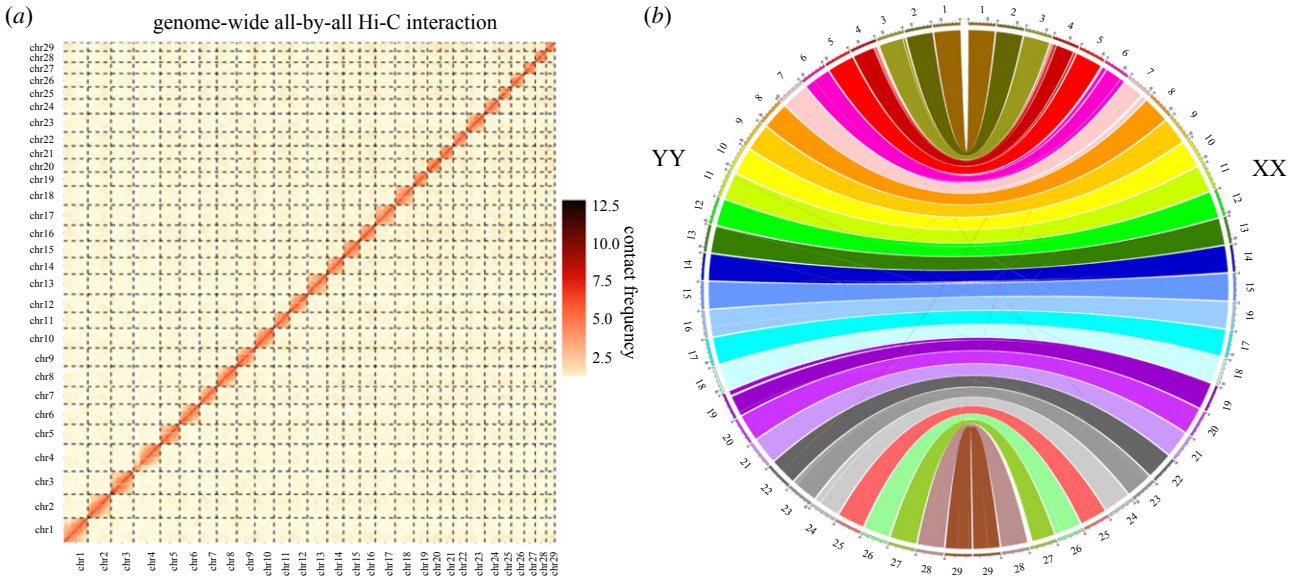

Figure 1. The YY genome assembly of the southern catfish. (*a*) Hi-C interactions among 29 chromosomes in YY genome. Strong interactions are indicated in dark red and weak interactions in yellow. (*b*) Circos graph of whole-genome synteny analysis for YY genome and previously reported XX genome [13]. Lines linking two chromosomes indicate the location of homologues. (Online version in colour.)

Chr24 of the YY genome harbouring SNPs significantly associated with phenotypic sex (figure 2*a*,*b*). This result confirmed that Chr24 is the sex chromosome of the southern catfish. The 71 protein-coding genes located in this SDR are listed in the electronic supplementary material, table S18.

(c) Identification of candidate sex-determining gene

The lengths of Chr24 in the XX female genome [13] and YY male genome are approximately 21.33 Mb and approximately 21.95 Mb, representing the sizes of ChrX and ChrY, respectively (electronic supplementary material, table S12). Alignments of the Nanopore long reads from XX, XY and YY individuals and the Illumina short reads from F-pool and M-pool mapped to the YY genome revealed a Y-specific insertion (from 4.67 to 5.22 Mb) in the SDR (figure 2*c*,*d*). The synteny analysis of X and Y chromosomes confirmed this Y-specific insertion (electronic supplementary material, figure S1). Annotation results showed that this insertion contained a single gene which is a duplicate copy (*amhr2y*) (from 4 771 130 bp to 4 780 169 bp) of the autosomal *amhr2* located on Chr8 (electronic supplementary material, table S18). Both *amhr2y* and *amhr2* have 11 exons. The lengths of *amhr2y* and *amhr2* coding region are 9.04 kb and 5.84 kb, respectively (figure 2*e*). Alignment of *amhr2y* and *amhr2* genomic sequences revealed that the nucleotide identity between exon sequences of the two genes ranges from 70.2% to 92.2% (average: 81.4%), while little sequence similarity was found between the 5 kb upstream sequences and introns of the two genes (figure 2*f*,*g*; electronic supplementary material, figure S2). Predicted proteins contain 496 amino acids (aa) for *amhr2* and 492 aa for *amhr2y*, sharing 70.6% identity and 78.6% similarity. Both proteins have a complete transmembrane domain, N-terminal intracellular protein kinase domain and 10 conserved cysteines at the C-terminus (electronic supplementary material, figure S3).

To confirm the male specificity of this Y-specific duplication, we designed five pairs of primers specific for *amhr2y* and one pair of primer common for *amhr2y* and *amhr2* and genotyped 25 female (XX), 25 male (XY) and 25 supermale (YY) individuals. The results showed that the bands of *amhr2* were present in all individuals, while the presence/absence of the bands of *amhr2y* were perfectly correlated with the phenotypic sex. The *amhr2y* fragment was observed in the 25 males and 25 supermales but not in the 25 females (electronic supplementary material, figure S4). These results indicated that *amhr2y* is a Y specific gene.

(d) Expression of *amhr2y* and *amhr2* in gonads at different developmental stages

To characterize the temporal and spatial expression of these two duplicates, RT-PCR, RNA-seq and FISH were performed. RT-PCR showed that among the 12 tissues examined, both *amhr2y* and *amhr2* were expressed exclusively in gonads. *Amhr2* was predominantly expressed in testis and expressed at a low level in the ovary/*Amhr2y* was expressed only in the testis, not in the ovary (figure 3*a*). Transcriptome analysis from gonads at three developmental stages revealed that *amhr2* was expressed in both male and female gonads, with significantly higher expression in male than in female gonads at all stages examined. By contrast, expression of *amhr2y* was detected only in male gonads, with expression peaked at 120 dah, and decreased at 360 dah and to very low level at 540 dah (figure 3*b*; electronic supplementary material, table S19). Early expressions and cellular localizations of *amhr2* and *amhr2y* were then characterized by FISH performed on male and female gonads sampled at 5, 10, 30 and 120 dah (figure 3*c*; electronic supplementary material, figure S5). The results showed that *amhr2y* and *amhr2* were expressed in both somatic cells and germ cells of male gonads at all these stages, and the expression of *amhr2y* was much stronger than that of *amhr2* at the early two stages (5 dah and 10 dah). By contrast, no expression of these two genes was detected in female gonads at any of the stages.

(e) Disruption of *amhr2y* by CRISPR/Cas9 caused male-to-female sex reversal

To gain functional support of *amhr2y* as an SD gene of southern catfish, we examined the gonadal development of the XY

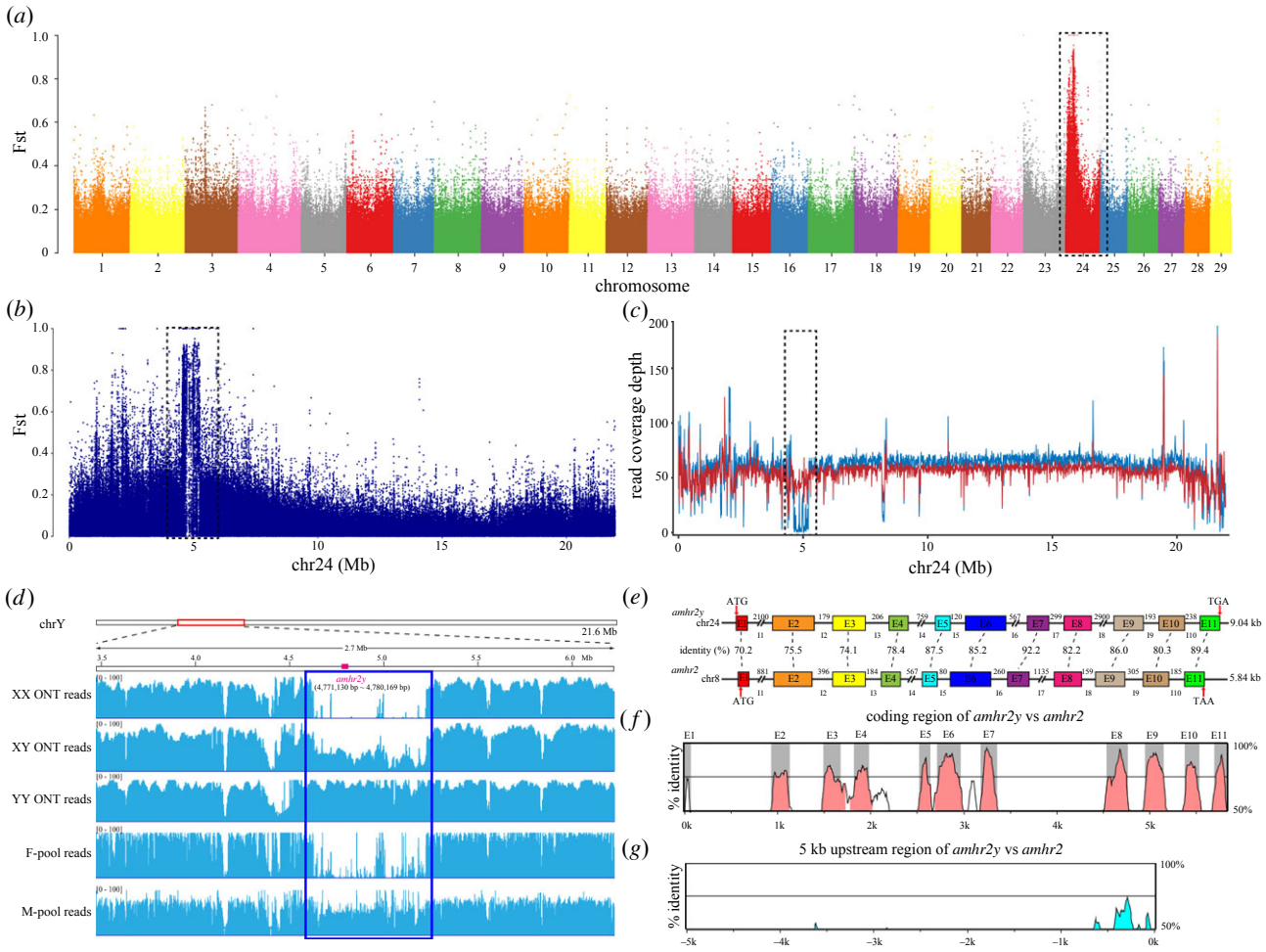

Figure 2. Identification of SDR, Y-specific insertion and candidate SD gene *amhr2y*. (*a*,*b*) Manhattan plot for SNP results associated with sex in each chromosome and Chr24 only, respectively. Chr24 showed a concentration of SNPs significantly associated with sex phenotype. (*c*) Coverage depth of male and female pool-seq reads on Chr24. A window is considered Y-specific if it is covered by few mapped female reads and by male reads at a depth close to half of the genome average. (*d*) Alignments of the ONT reads from XX, XY and YY individuals and the Illumina reads from F-pool and M-pool to the YY genome revealed a Y-specific insertion (marked by a blue box) in the SDR which contained a male-specific duplicate of *amhr2* (named *amhr2y*). ONT, Oxford Nanopore Technologies. (*e*–*g*) Gene structure and sequence identity between *amhr2* and *amhr2y*. Exons are represented by boxes with shared percentage identity indicated and introns are represented by black lines. The start codon of each gene is positioned at 0 kb (*f*,*g*). (Online version in colour.)

loss-of-function mutants for *amhr2y* using CRISPR/Cas9 system in the F0 generation. The target was selected in the exon 2 of *amhr2y*. The guide RNA (gRNA) contained *Tsp*RI sites adjacent to the protospacer adjacent motif (PAM) sequence for mutation analysis. Complete digestion with *Tsp*RI produced two fragments (135 and 235 bp) in the control group, whereas an intact DNA fragment (approx. 370 bp) was observed in embryos injected with both Cas9 mRNA and gRNA. The mutations, including in-frame and frame-shift deletions, were further confirmed by Sanger sequencing (figure 4*a*). Off-target mutations in autosomal *amhr2* were not detected (electronic supplementary material, figure S6). Fertilized eggs microinjected with *amhr2y* gRNA and Cas9 mRNA were hatched, and larvae were cultured to 90 dah.

When screening *amhr2y* mutated fish, *Tsp*RI digestion of the DNA fragments spanning the *amhr2y* target site from the 12 individuals indicated the mutation rate from 14% to 81% occurring in the target region (figure 4*b*). Further, genetic sex was confirmed by sex-specific marker and phenotypic sex was determined by histological observation of gonads. Amplification results revealed that all of these *amhr2y* mutant individuals were XY fish as expected (figure 4*c*). Histological examination of gonads revealed that some of the mutated F0 fish with high-rate *amhr2y* knockout (greater than 55%) showed male to female sex reversal (figure 4*d*). The mutation types of the sex-reversed individuals were analysed by Sanger sequencing of the uncleaved bands after *Tsp*RI digestion, and 86% to 100% of mutations were frame shifts (electronic supplementary material, figure S7). Gonadal sex differentiation in the sex-reversed XY fish was characterized by the formation of the OC and the appearance of primary oocytes. These gonads were topologically indistinguishable from the control XX ovary, but obviously different from the control XY testis. Additionally, the gonads of those mutant fish with low knockout rates still developed normal testis (figure 4*d*).

A total of 124 injected F0 individuals survived and were analysed. Sixty-three (51%) individuals were genetically male (XY). Mutations by CRISPR/Cas9 were detected in 29 (46%) of these genetic males. Nine (31%) of these individuals with high mutation rate were induced to male-to-female sex reversal (figure 4*e*; electronic supplementary material, figure S8).

(f) Syntenic and phylogenetic analysis of *amhr2* genes

Male-specific duplication of *amhr2* has also been identified in yellow perch [35], ayu [36], common seadragon and its

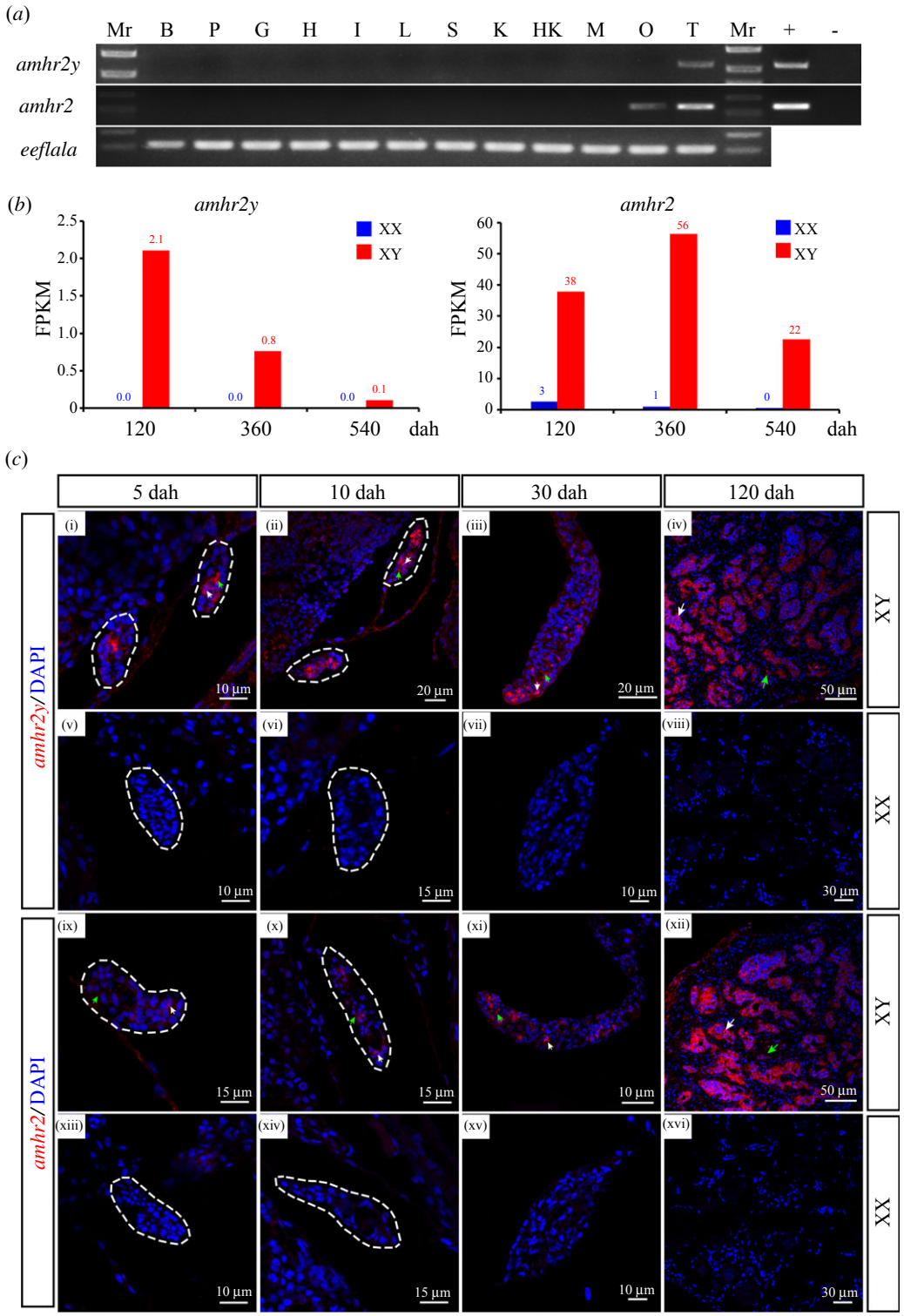

Figure 3. Expression analyses of *amhr2* and *amhr2y* in southern catfish. (*a*) Tissue distribution at 120 dah (day after hatching) by RT-PCR. B, brain; P, pituitary; G, gill; H, heart; I, intestine; L, liver; S, spleen; K, kidney; HK, head kidney; M, muscle; O, ovary; T, testis; Mr, DNA marker DL2000; +, positive control; -, negative control. *eef1a1a* as internal control. (*b*) Gonadal transcriptome analysis of southern catfish at 120, 360 and 540 dah. FPKM, fragments per kilobase of exon per million reads mapped. (*c*) FISH of male and female gonads in southern catfish at 5, 10, 30 and 120 dah. White arrows indicate the germ cells and green arrows indicate the somatic cells. (Online version in colour.)

closely related species alligator pipefish [37]. To obtain the traces of Y-specific duplication of *amhr2*, a syntenic map was generated for *amhr2*, *amhr2y* and their adjacent genes (electronic supplementary material, figure S9a). Fifteen teleost species were used for syntenic analysis with spotted gar as an outgroup. Genes located upstream (i.e. *cdca7*, *sp1* and *sp7*) and downstream (i.e. *pcbp2*, *map3k12* and *myg1*) of *amhr2* on spotted gar LG4 showed conserved synteny in all teleosts except for the four catfishes included in the analysis, indicating that transposition of *amhr2* occurred in the common ancestor of these four catfishes. Genes located upstream (i.e. *galcb*) and downstream (i.e. *ak7b*, *vrk1* and *bcl11b*) of *amhr2* on Chr8 showed conserved synteny in all four catfish species, indicating that Chr8 is the conserved location of the southern catfish ancestral *amhr2*. However, there is no obvious synteny between *amhr2*-containing region in Chr8 and *amhr2y*-containing region in Chr24 of southern catfish. In addition, the *amhr2y*-containing region in southern catfish Chr24, yellow perch Chr9, ayu

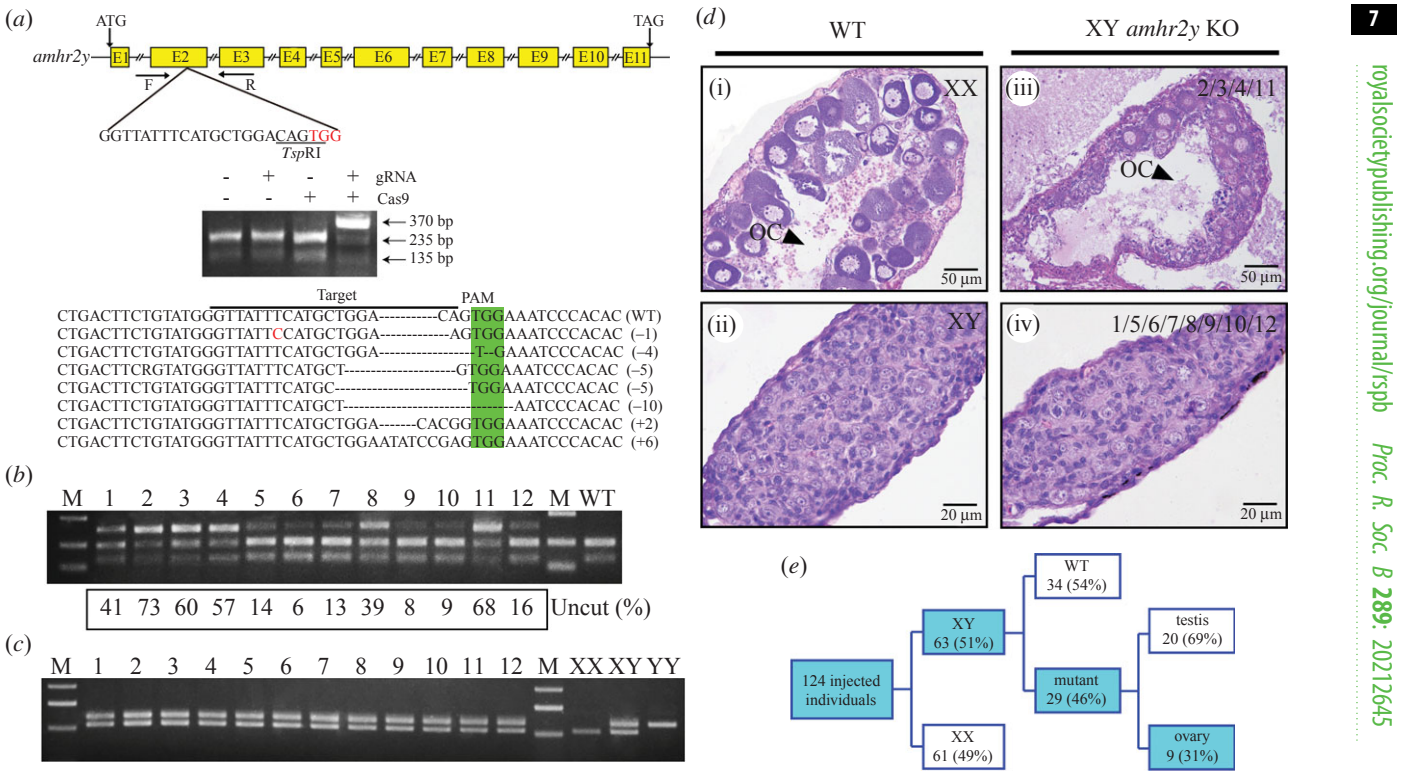

Figure 4. Mutation of *amhr2y* by CRISPR/Cas9 resulted in male to female sex reversal in F0 XY fish. (*a*) Gene structure of *amhr2y* showing the target site and the *Tsp*RI restriction site. Mutation sequences from the uncleaved bands are listed. The PAM is marked in light green. Numbers to the right of the sequences indicate the loss (–) or gain (+) of bases for each allele. WT, wild-type. (*b*) Twelve *amhr2y* KO XY F0 mutants detected by restriction enzyme *Tsp*RI digestion. The percentage of un-cleaved (i.e. mutant) DNA is shown for each individual. (*c*) Genotype of the *amhr2y* mutant fish was confirmed with a sex-specific marker (marker-8) reported previously [12]. (*d*) Gonadal histology of a representative control XX female (i), control XY male (ii) and *amhr2y* KO XY individuals (iii, iv). Some of the *amhr2y* KO XY individual (iii) developed ovaries with oocytes and an ovary cavity, indistinguishable from the ovary of the control females (i). OC, ovarian cavity. (*e*). Statistics of genotypic and phenotypic sex of all injected individuals. (Online version in colour.)

scaffold C, common seadragon Chr4 and alligator pipefish Chr21 did not show any synteny with each other. These results suggest that *amhr2y* evolved from independent transposition events in each lineage.

Further phylogenetic analysis of teleost Amhr2, including those species with male-specific *amhr2* duplications, was performed to test whether they have a shared origin (electronic supplementary material, figure S9b). In the protein phylogenic tree, Amhr2y of common seadragon and alligator pipefish clustered together as a sister clade to the two species' canonical Amhr2 with significant bootstrap values, indicating that a lineage-specific duplication event has taken place in the common ancestor of these two species. However, the Amhr2y of southern catfish, yellow perch and ayu clustered with their own paralogue, suggesting independent lineage-specific duplication events in these species.

4. Discussion

Teleost X and Y chromosomes are highly similar, making it difficult to separately assemble the sequences of X and Y chromosomes. The YY supermale provided sequencing templates of the Y chromosome without contamination by homologous X chromosome sequences that could confound assembly [38]. To date, whole Y chromosome sequences have been generated for only two teleosts: the channel catfish [39] and three-spine stickleback [40]. The sex chromosomes of southern catfish are homomorphic [11]. We previously generated a well-assembled X chromosome from an XX female [13]. In the present study, YY males were obtained by crossing normal XY males with experimentally sex-reversed XY neofemales. Through a combination of Illumina, Nanopore and Hi-C sequencing, we were able to assemble a highly accurate and complete Y chromosome using the YY male as the sequencing template. In addition, sequencing of XY fish provided a good control when analysing read coverage between the X and Y chromosomes. Taken together, genome sequencing and assembly of XX, XY and YY provided a unique system for comparative genome analysis of the X and Y sequences.

The location of male-specific markers revealed that Chr24 is the sex chromosome of southern catfish. In this study, we identified a small SDR of 2.38 Mb on this sex chromosome using a male versus female pooled gDNA whole-genome sequencing strategy which has been often used to explore SDR in vertebrates [41–43]. The Y-specific region of approximately 500 kb was further determined by examining the relative depth of coverage when aligning male versus female resequencing reads and ONT reads against the YY genome. This Y-specific region contains a duplicate (*amhr2y*) of the autosomal (Chr8) *amhr2*. Results from PCR-based genotyping in a large number of individuals demonstrate a strong and significant association of *amhr2y* with male phenotype. Therefore, the PCR assay developed based on *amhr2y* has the potential to improve fisheries management and conservation in southern catfish, as it can be used as a universal marker for genetic sex identification of individuals.

amhr2 has previously been characterized as an SD gene in some pufferfishes (*Takifugu rubripes*, *T. pardalis* and *T. poecilonotus*) [44,45]. However, in contrast with pufferfishes,

in which the differentiation of X and Y chromosomes is extremely limited and originated from an allelic diversification process, the southern catfish *amhr2y* sequence is quite divergent from its *amhr2* autosomal counterpart. Specifically, *amhr2y* shows only 81.0% identity with *amhr2* in the coding region. Moreover, the sequence identity of the promotor and intron regions is extremely low. These results suggest that the duplication is likely to be ancient. The accumulation of their sequence differences may contribute to functional divergence.

Except for its Y chromosome-specific location, *amhr2y* could be considered a candidate SD gene of southern catfish also because of its male-specific expression profile. According to the RT-PCR and transcriptome data, *amhr2y* mRNA was detected only in male gonads. Due to the difficulty in sampling gonad tissues at early stages, the expression analysis of *amhr2y* was only conducted in three pairs of gonadal transcriptomes from 120, 360 and 540 dah. FISH was performed to analyse its expression and cellular localization in early developmental stages of male and female gonads. The results showed that *amhr2y* was only expressed in male gonads, with high expression at 5 dah, the critical sex-determination period, clearly preceding the first signs of morphological differentiation of ovaries and testes at approximately 10 dah [14].

The effective way to understand the function of a gene is to perform gain/loss of function studies and to characterize the resulting biological effects. In the present study, mutation of *amhr2y* using CRISPR/Cas9 resulted in complete gonadal sex reversal of XY mutants, demonstrating that *amhr2y* is necessary for testicular differentiation. Together with the Y-specific location and male-specific expression profile, these independent lines of evidence provide strong support for the idea that *amhr2y* is the SD gene in southern catfish.

Amh/Amhr2 signalling is critical for sex determination in various species of teleosts [5]. However, little is known about the difference in function between *amhr2y* and autosomal *amhr2*. Based on genetic evidence and expression patterns, several models were proposed to explain how *amhr2y* determined the genetic sex in ayu [36]. These models cover three aspects: (i) Amhr2y and autosomal Amhr2 bind to the same ligand, but the downstream signalling pathway may differ between them because of the mutation in the kinase domain of Amhr2y; (ii) Amhr2y and Amhr2 bind to different ligands, and testicular development might be initiated by Amhr2y and the specific ligand complex; (iii) Amhr2y and Amhr2 bind to the same ligand and have exactly the same function, the gene dosage of Amhr2 might be critical for sex determination. In southern catfish, because of the relatively high level of divergence between Amhr2 and Amhr2y sequences, it is tempting to hypothesize that the two proteins could have diverged in their function. For instance, due to divergence in the N-terminal extracellular domain, they may have a different affinity for their canonical Amh ligand or Amhr2y may even have the ability to bind other ligands. One possible ligand is Gsdf which was demonstrated to be critical for testicular differentiation in rainbow trout [46], medaka [47] and tilapia [48]. The rates/patterns of downstream phosphorylation may also differ due to divergence in the C-terminal domain. On the other hand, although *amhr2y* and *amhr2* are expressed in the same cell types of the XY gonad, the higher expression of *amhr2y* in the early stages (5 and 10 dah) may also be the reason why it becomes an SD gene. Further functional studies would be required to unravel the upstream and downstream signalling pathways of Amhr2 and Amhr2y in southern catfish to better understand the mechanisms leading to the novel function of *amhr2y* as the SD gene in this species.

Gene duplication/insertion is one of the most important mechanisms for the origin of new SD genes, which could suppress recombination more readily than allelic diversification [49,50]. Male-specific duplication of *amhr2* has been identified in four species of teleosts, namely yellow perch [35], ayu [36], common seadragon and alligator pipefish [37]. Our present study provided a new example of an *amhr2* duplication event. Syntenic and phylogenetic analyses revealed that these male-specific *amhr2* duplications are independent within each lineage. In addition, other members of the TGF-β pathway, such as *amh*, *gsdf*, *gdf6* and *bmpr1b*, are recruited as SD genes in different fishes (reviewed in [5]). The repeated, independent recruitment of the same gene for sex-determination supports the 'limited options' hypothesis for the evolution of genetic SD mechanisms and make TGF-β, especially AMH/AMHR2, pathway members the most frequently recruited SD genes identified in fishes so far [51]. However, the evolutionary pressures that lead to recurrent convergent evolution toward using TGF-β signalling pathways to control sex determination in fish remain elusive.

Ethics. Animal experiments in this study were conducted in accordance with the regulations of the Guide for Care and Use of Laboratory Animals and were approved by the Committee of Laboratory Animal Experimentation at Southwest University.

Data accessibility. The Whole Genome Shotgun (WGS) data have been deposited at the National Genomics Data Center (https://ngdc.cncb.ac.cn/gwh/) under BioProject PRJCA002907 with accession GWHAOSD00000000 and GWHBHES00000000. The raw genome and RNA sequencing data have been deposited in the Sequence Read Archive (SRA) database (SRR11929421-SRR11929440) under BioProject PRJNA631665. The data are provided in the electronic supplementary material [52].

Authors' contributions. S.Z.: funding acquisition, investigation, methodology, validation, writing—original draft and writing—review and editing; W.T.: data curation, formal analysis and writing—review and editing; H.Y.: validation and visualization; T.D.K.: methodology and writing—review and editing; Z.W.: supervision; Z.P.: supervision; L.J.: supervision; D.P.: supervision; Y.Z.: conceptualization, formal analysis and supervision; D.W.: conceptualization, formal analysis, funding acquisition, writing—original draft and writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by grants from National Key Research and Development Program of China (2018YFD0900202), National Natural Science Foundation of China (31861123001 and 31630082), Chongqing Science and Technology Bureau (cstc2021ycjh-bgzxm0024 and CQYC201903173), China Postdoctoral Science Foundation (2021M692680) and Chongqing Postdoctoral Science Foundation (cstc2021jcyj-bsh0061).

References

1. Kikuchi K, Hamaguchi S. 2013 Novel sex-determining genes in fish and sex chromosome evolution. *Dev. Dyn.* **242**, 339–353. (doi:10.1002/dvdy.23927)

2. Herpin A, Schartl M. 2015 Plasticity of gene-regulatory networks controlling sex determination:

of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep.* **16**, 1260–1274. (doi:10.15252/embr.201540667)

3. Heule C, Salzburger W, Bohne A. 2014 Genetics of sexual development: an evolutionary playground for fish. *Genetics* **196**, 579–591. (doi:10.1534/genetics.114.161158)
4. Pan QW *et al.* 2016 Vertebrate sex-determining genes play musical chairs. *C. R. Biol.* **339**, 258–262. (doi:10.1016/j.crvi.2016.05.010)
5. Pan QW *et al.* 2021 Evolution of master sex determiners: TGF-beta signalling pathways at regulatory crossroads. *Phil. Trans. R. Soc. B* **376**, 20200091. (doi:10.1098/rstb.2020.0091)
6. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991 Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121. (doi:10.1038/351117a0)
7. Smith CA *et al.* 2009 The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* **461**, 267–271. (doi:10.1038/nature08298)
8. Schartl M, Schmid M, Nanda I. 2016 Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* **125**, 553–571. (doi:10.1007/s00412-015-0569-y)
9. Devlin RH, Nagahama Y. 2002 Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364. (doi:10.1016/s0044-8486(02)00057-1)
10. Xie XJ, Long TC, Cao ZD. 1994 Studies on the composition and growth in the reproductive population of *Silurus meridionalis*. *J. Southwest China Norm. Univ.* **19**, 71–78. (doi:10.13718/j.cnki.xsxb.1994.01.015)
11. Hong YH, Zhou T. 1983 Karyotypic study of two species of Chinese catfishes (Siluridae). *J. Wuhan Univers.* **3**, 108–130. (doi:10.14188/j.1671-8836.1983.03.008)
12. Zheng SQ, Wang XS, Zhang S, Long J, Tao WJ, Li MH, Wang DS. 2020 Screening and characterization of sex-linked DNA markers and marker-assisted selection in the southern catfish (*Silurus meridionalis*). *Aquaculture* **517**, 10. (doi:10.1016/j.aquaculture.2019.734783)
13. Zheng SQ *et al.* 2021 Chromosome-level assembly of southern catfish (*Silurus meridionalis*) provides insights into visual adaptation to nocturnal and benthic lifestyles. *Mol. Ecol. Resour.* **21**, 1575–1592. (doi:10.1111/1755-0998.13338)
14. Dong RR, Yang SJ, Feng RJ, Fang LL, Sun YL, Zhang YG, Xie XJ, Wang DS. 2014 Complete feminization of catfish by feeding *Limnodilus*, an annelid worm collected in contaminated streams. *Environ. Res.* **133**, 371–379. (doi:10.1016/j.envres.2014.05.018)
15. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
16. Rao SS *et al.* 2014 A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680. (doi:10.1016/j.cell.2014.11.021)
17. Jain M *et al.* 2018 Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345. (doi:10.1038/nbt.4060)
18. Walker BJ *et al.* 2014 Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963. (doi:10.1371/journal.pone.0112963)
19. Durand NC *et al.* 2016 Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98. (doi:10.1016/j.cels.2016.07.002)
20. Dudchenko O *et al.* 2017 De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95. (doi:10.1126/science.aal3327)
21. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
22. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
23. Li H *et al.* 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
24. McKenna A *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
25. Li H. 2018 Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. (doi:10.1093/bioinformatics/bty191)
26. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014 deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191. (doi:10.1093/nar/gku365)
27. Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192. (doi:10.1093/bib/bbs017)
28. Keilwagen J, Hartung F, Grau J. 2019 GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177. (doi:10.1007/978-1-4939-9173-0_9)
29. Dobin A *et al.* 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
30. Li B, Dewey CN. 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323. (doi:10.1186/1471-2105-12-323)
31. Adolfi MC, Herpin A, Regensburger M, Sacquegno J, Waxman JS, Schartl M. 2016 Retinoic acid and meiosis induction in adult versus embryonic gonads of medaka. *Sci. Rep.* **6**, 34281. (doi:10.1038/srep34281)
32. Li MH *et al.* 2014 Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics* **197**, 591–599. (doi:10.1534/genetics.114.163667)
33. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. (doi:10.1093/molbev/mst197)
34. Nguyen NTT, Vincens P, Roest Crollius H, Louis A. 2018 Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Res.* **46**, D816–D822. (doi:10.1093/nar/gkx1003)
35. Feron R *et al.* 2020 Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Mol. Ecol. Resour.* **20**, 531–543. (doi:10.1111/1755-0998.13133)
36. Nakamoto M *et al.* 2021 A Y-linked anti-Mullerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS Genet.* **17**, e1009705. (doi:10.1371/journal.pgen.1009705)
37. Qu M *et al.* 2021 Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Sci. Adv.* **7**, eabg5196. (doi:10.1126/sciadv.abg5196)
38. Tomaszkiewicz M *et al.* 2016 A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. *Genome Res.* **26**, 530–540. (doi:10.1101/gr.199448.115)
39. Bao L *et al.* 2019 The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *BMC Biol.* **17**, 6. (doi:10.1186/s12915-019-0627-7)
40. Peichel CL *et al.* 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* **21**, 177. (doi:10.1186/s13059-020-02097-x)
41. Gammerdinger WJ, Conte MA, Acquah EA, Roberts RB, Kocher TD. 2014 Structure and decay of a proto-Y region in tilapia, *Oreochromis niloticus*. *BMC Genomics* **15**, 975. (doi:10.1186/1471-2164-15-975)
42. Gammerdinger WJ, Conte MA, Baroiller JF, D'Cotta H, Kocher TD. 2016 Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. *BMC Genomics* **17**, 808. (doi:10.1186/s12864-016-3163-7)
43. Gammerdinger WJ, Conte MA, Sandkam BA, Ziegelbecker A, Koblmuller S, Kocher TD. 2018 Novel sex chromosomes in 3 cichlid fishes from Lake Tanganyika. *J. Hered.* **109**, 489–500. (doi:10.1093/jhered/esy003)
44. Kamiya T *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. (doi:10.1371/journal.pgen.1002798)
45. Ieda R *et al.* 2018 Identification of the sex-determining locus in grass puffer (*Takifugu niphobles*) provides evidence for sex-chromosome turnover in a subset of *Takifugu* species. *PLoS*

ONE **13**, e0190635. (doi:10.1371/journal.pone.0190635)

46. Sawatari E, Shikina S, Takeuchi T, Yoshizaki G. 2007 A novel transforming growth factor-beta superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). *Dev. Biol.* **301**, 266–275. (doi:10.1016/j.ydbio.2006.10.001)
47. Zhang X *et al.* 2016 Autosomal *gsdf* acts as a male sex initiator in the fish medaka. *Sci. Rep.* **6**, 19738. (doi:10.1038/srep19738)
48. Kaneko H, Ijiri S, Kobayashi T, Izumi H, Kuramochi Y, Wang DS, Mizuno S, Nagahama Y. 2015 Gonadal soma-derived factor (*gsdf*), a TGF-beta superfamily gene, induces testis differentiation in the teleost fish *Oreochromis niloticus*. *Mol. Cell Endocrinol.* **415**, 87–99. (doi:10.1016/j.mce.2015.08.008)
49. Herpin A, Schartl M. 2009 Molecular mechanisms of sex determination and evolution of the Y-chromosome: insights from the medakafish (*Oryzias latipes*). *Mol. Cell Endocrinol.* **306**, 51–58. (doi:10.1016/j.mce.2009.02.004)
50. Kondo M, Nanda I, Schmid M, Schartl M. 2009 Sex determination and sex chromosome evolution: insights from medaka. *Sex Dev.* **3**, 88–98. (doi:10.1159/000223074)
51. Marshall Graves JA, Peichel CL. 2010 Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol.* **11**, 205. (doi:10.1186/gb-2010-11-4-205)
52. Zheng S *et al*. 2022 Identification of sex chromosome and sex-determining gene of southern catfish (*Silurus meridionalis*) based on XX, XY and YY genome sequencing. FigShare.