GENETICS

Transposons drove the evolution of a male sex-determining gene from a female gene with deep introgression across cichlids

Le Wang[1]†, Fei Sun[1]†, Zituo Yang[1,2]†, May Lee[1], Joey Wong[1], Shadame Yeo[1], Yanfei Wen[1], Axel Meyer[3,4,5,6]*, Francesc Piferrer[7]*, Manfred Schartl[8,9,10]*, Gen Hua Yue[1,11]*

Sex-determination mechanisms and their master regulators are diverse among vertebrates. Components of downstream regulatory pathways can sometimes be co-opted as new upstream triggers. Here, we report a notable finding of cross-pathway recruitment: A gene from the female differentiation cascade was repurposed as a male master sex-determining (MSD) gene. We identify *figlaY*, a Y-linked truncated duplicate of the female-associated *figla*, as the MSD gene in Mozambique tilapia. *figlaY* originated through duplication and translocation in an ancestral cichlid lineage and subsequently introgressed into multiple lineages. Transposable elements facilitated *figlaY*'s emergence by mediating duplication, truncation, and regulatory rewiring. Functionally, FIGLAY suppresses the *zp2* promoter activated by the FIGLA/E12 heterodimer, acting as a dominant-negative regulator that inhibits ovarian differentiation and promotes testis development. This discovery reveals an unprecedented evolutionary route from female to male sex determination, i.e., a "genetic defection" from an initially female pathway. This previously unknown mechanism expands the vertebrate sex-determination toolkit.

INTRODUCTION

Sex determination in vertebrates is governed by diverse mechanisms that involve the interplay of genetic, epigenetic, environmental, ecological, and physiological factors (*1*). Among vertebrates, teleost fishes are particularly notable for their extraordinary variation, not only in the sexual systems that they exhibit (*2*) but also in the identity and evolutionary origins of their master sex-determining (MSD) genes, that can differ strongly even among closely related species (*1*–*3*). Despite this diversity, only around 20 different MSD genes have been identified across vertebrates, many of which have evolved independently and repeatedly (*4*). Most of these MSD genes are linked to the transforming growth factor–β signaling pathway, while others repeatedly use transcription factors such as *dmrt1* and *sox3* or steroidogenic enzyme genes such as *hsd17b1* (*4*, *5*). These findings support the prevailing notion that a limited set of "usual suspects" repeatedly emerge as evolutionary solutions in the establishment of genotypic sex determination.

However, growing evidence suggests that the evolution of MSD genes and the turnover of sex chromosomes may be even more flexible than previously thought. Several hypotheses propose that genes already integrated within regulatory networks of sex determination and differentiation may become MSD genes following genetic changes that rewire transcriptional regulation (*6*, *7*). Gene duplication and subsequent specialization by neo- and subfunctionalization represent another major molecular evolutionary mechanism for the origin of novel MSD genes (*8*). In addition, transposable elements (TEs) are increasingly recognized as powerful drivers of MSD gene emergence. TEs can mediate gene duplication and truncation, promote neofunctionalization, and generate allelic and expression diversity (*9*, *10*). Through cis-regulatory modulation and epigenetic silencing, these activities accelerate genomic changes underlying evolutionary innovations and create opportunities for the recruitment of novel MSD genes (*11*–*15*). Collectively, these mechanisms indicate that the pool of MSD candidates may not be restricted to a handful of recurrent genes but instead reflects both the intrinsic plasticity of sex-determination networks and the dynamic activity of TEs. This implies that many MSD genes remain undiscovered, and their identification will be critical for advancing our understanding of sex chromosome evolution and the diversity of sex-determination systems (*16*, *17*).

Species commonly referred to as "tilapia," encompassing nearly 100 African and Middle Eastern cichlids (*18*), provide an exceptional model for investigating these questions. They are also among the world's most important aquaculture species. Four sex-determining loci have been identified in tilapia—on LG1 (XY), LG3 (ZW), LG14 (XY), and LG23 (XY) (*19*). These loci have been extensively introgressed across lineages due to both natural hybridization and artificial breeding practices (*20*–*24*). Despite intensive research, only one MSD gene, *amhY* on chromosome 23, has been identified and functionally validated in Nile tilapia (*Oreochromis niloticus*) (*25*). Yet, the precise MSD genes controlling the LG1 and LG3 systems remain unknown. Recently, a *figla*-like gene was found to be male specific in tilapia species with the LG1 system (*20*), although whether it represents the true MSD gene at this locus or is merely functionally associated with male development remains unclear.

[1]Temasek Life Sciences Laboratory, Singapore 117604, Singapore. [2]College of Marine Sciences, South China Agricultural University, Guangzhou 510642, China. [3]Department of Biology, University of Konstanz, 78457 Konstanz, Germany. [4]Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA. [5]CAS Key Laboratory of Tropical Marine Bio-Resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China. [6]MOE Key Laboratory of Freshwater Fish Reproduction and Development, College of Fisheries, School of Life Sciences, Integrative Science Center of Germplasm Creation in Western China (Chongqing) Science City, Southwest University, Chongqing 400715, China. [7]Institute of Marine Sciences (ICM), Spanish National Research Council (CSIC), 08003 Barcelona, Spain. [8]Institute of Pathology, University of Wuerzburg, 97074 Wuerzburg, Germany. [9]The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, TX 78666, USA. [10]Research Department for Limnology, University of Innsbruck, 5310 Mondsee, Austria. [11]Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore.
*Corresponding author. Email: axel.meyer@uni-konstanz.de (A.M.); piferrer@icm.csic.es (F.P.); phch1@biozentrum.uni-wuerzburg.de (M.S.); genhua@tll.org.sg (G.H.Y.)
†These authors contributed equally to this work.

Here, we report the identification and validation of *figlaY* as the MSD gene on LG1 in a Singapore strain of Mozambique tilapia (*Oreochromis mossambicus*), originally derived from South Africa. *figlaY* is a Y-specific paralog of the transcription factor *figla* (factor in the germline 1 alpha), a conserved gene that plays a key role in female germ cell development and is distinct from all previously known sex-determining genes. We demonstrate that *figlaY* originated through gene duplication and translocation, followed by neofunctionalization, in an ancestral cichlid lineage and was subsequently introgressed into multiple lineages, including tilapia and Lake Tanganyikan tribes. Furthermore, we provide evidence that TEs played a central role in the origin of this previously unidentified MSD gene and propose a mechanism by which a paralog of a typically female-associated gene was co-opted into the male sex-determination pathway.

RESULTS AND DISCUSSION

An additional gene in the Y-specific sex-determining region of LG1

Previous studies have shown that an MSD gene located on LG1 in the Nile tilapia (*O. niloticus*) operates within an XY sex-determination system. Also, a Singapore strain of Mozambique tilapia (*O. mossambicus*), derived from a wild population in South Africa, follows the LG1 sex-determination system (*26*). To identify the MSD gene, we sequenced and assembled the genomes of an XX female and a YY male of *O. mossambicus*, using Nanopore and PacBio technologies. The XX and YY genome assemblies have total lengths of 1026.9 and 1001.6 Mb, with contig N50 sizes of 2.073 and 2.491 Mb, and scaffold N50 sizes of 40.585 and 40.730 Mb, respectively (table S1). Each genome contains 22 haploid pseudochromosomes (fig. S1 and table S2). The Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness scores for the XX and YY assemblies are 98.3 and 98.7%, respectively, with 29,428 and 28,656 predicted protein coding genes (table S1). Repetitive sequences account for 38.1 and 37.1% of the XX and YY genomes, respectively (table S3). The genome size, repetitive content, and gene predictions are comparable to previously reported tilapia genome sequences (*27*). The slightly greater length (+2.5%) and gene content (+2.7%) in the XX assembly are likely attributable to the higher error rate of Nanopore sequencing, leading to increased haplotig divergence and incomplete purging.

To identify the sex-determining locus, we constructed a linkage map comprising 6779 single-nucleotide polymorphisms (SNPs), across 22 linkage groups (LGs) (table S4). Quantitative trait locus (QTL) mapping pinpoints the sex-determining locus to a ~3.5-Mb interval (LG1, 21.9 to 25.4 Mb) in the YY genome (Fig. 1A and fig. S2). SNPs within this QTL show fixed differences between sexes, with 124 homozygous and 119 heterozygous individuals in a mapping family of 243 individuals. Further fine mapping in a mass-cross population of 960 individuals using SNP and Insertion-deletion (InDel) markers refined the locus to a ~950-kb region (LG1, 23.3 to 24.2 Mb) (fig. S3). An InDel sex marker developed within this region was used for genotyping both the mapping family and a broodstock population of 1169 individuals, showing tight linkage with phenotypic sex (Fig. 1, B and C). These data provide strong evidence

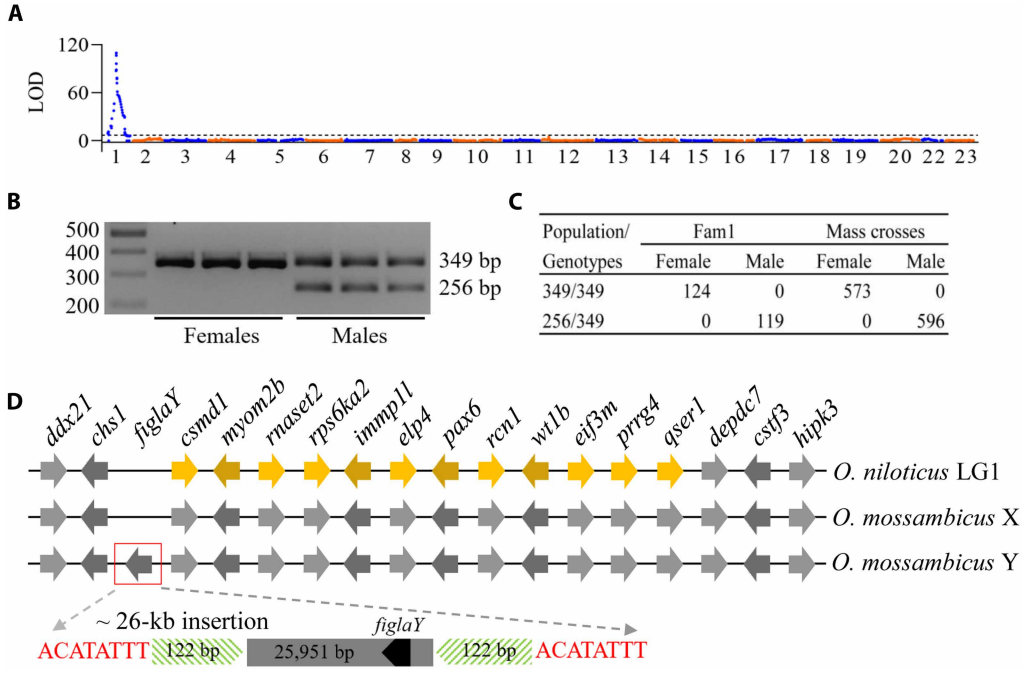

Fig. 1. The SDR of the LG1 sex-determination system in tilapia contains an additional gene in the Y-specific region. (**A**) QTL mapping identifies the sex-determining locus on LG1 in Mozambique tilapia. The genome-wide significance threshold (LOD) is indicated by a dotted line. The LGs on the *x* axis are numbered according to the *O. niloticus* assembly. (**B**) Development of an InDel-based sex marker in the sex-determining region (SDR) for polymerase chain reaction (PCR)–based genetic sex identification, which is tightly linked to the peak SNPs identified in the QTL analysis. (**C**) Genotypes at the sex marker are tightly linked with phenotypic sex across the entire mapping population. (**D**) Predicted protein-coding genes in the SDR of the X and Y chromosomes of Mozambique tilapia and the LG1 of Nile tilapia. The additional gene in the Y allele is embedded within a ~26-kb insertion, exhibiting hallmark features of a DNA retrotransposon, including core sequences flanked by an 8–base pair (bp) direct repeat and a 122-bp terminal inverted repeat. The position of the protein coding gene *figlaY* is indicated within the TE.

for a single-locus sex-determination mechanism in the studied population. The candidate region harbors 18 predicted protein coding genes on the Y chromosome and 17 in the corresponding region on the X chromosome in both Mozambique and Nile tilapia (Fig. 1D). The sex-determining region (SDR) in Mozambique tilapia overlaps with but is slightly larger than the reported SDR in Nile tilapia, which contains 12 genes (*28*). Sequence comparisons between the X and Y corresponding regions revealed that the additional gene in the Y region resides within a 26-kb male-specific insertion consistently present across our breeding population of Mozambique tilapia (fig. S4). This insertion exhibits features characteristic of a DNA retrotransposon, including flanking 122–base pair (bp) inverted and 8-bp direct repeats (Fig. 1D). The Y-specific gene within this region was identified as a homolog of the *factor in the germline alpha* (*figla*) and was therefore named *figlaY*. In contrast, the original *figla* gene is located on autosomal region of LG12.

figlaY shows male-biased expression

To identify the MSD gene(s) on the Y, we first analyzed genetic variants in the SDR by comparing X and Y allele sequences. We identified 12 nonsynonymous SNPs across seven genes, with five of these genes located in the SDR of Nile tilapia (Fig. 1D) (*28*). Among these, four genes show amino acid substitutions that alter biochemical properties, whereas one gene (*elp4*) contains a premature stop codon, retaining ~81% of its wild-type open reading frame (ORF) (table S5). Functional annotation revealed that only *WT1 transcription factor b* (*wt1b*) is known to play a role in gonadal development, whereas the others, as far as extensive literature search revealed, showed little evidence of involvement in sex determination or gonadal differentiation. However, knockout of *wt1b* in Nile tilapia did not affect sex phenotype (*29*).

Next, we examined the expression patterns of all 18 genes in the SDR. Histological analysis indicated that gonads in both XX and XY genotypes were at undifferentiated and differentiating stages at 20- and 30-days postfertilization (dpf), respectively (fig. S5). Transcriptome analysis of pooled gonads from XX and XY individuals revealed that *figlaY* consistently displayed the clearest male-biased differential expression (Fig. 2, A and B), whereas the remaining 17 genes showed no clear or consistent male-biased expression at 20 dpf and 30 dpf or in mature gonads (fig. S6). Even at an earlier stage (15 dpf), quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of pooled trunk samples showed that *figlaY* was the only gene among the 18 examined to exhibit notable male-biased expression (fig. S7). In contrast, the autosomal *figla* gene, located on LG12, showed female-biased expression in gonads before sexual maturation, although it maintained a basal level of expression in XY gonads (fig. S8). Notably, *figlaY* expression was restricted to the testis, whereas *figla* was detected in both the testis and ovary (Fig. 2C). In situ hybridization (ISH) analysis further demonstrated that *figlaY* was exclusively expressed in the supporting somatic cells surrounding germ cells in XY gonads at 20 dpf (Fig. 2D). In comparison, *figla* was also expressed in the supporting somatic cells surrounding germ cells in both XX and XY gonads, with higher expression in XX individuals (Fig. 2E). Collectively, the male-biased and gonad-specific expression patterns suggest that *figlaY* is a strong candidate MSD gene in Mozambique tilapia with the LG1 sex-determination system.

figlaY knockout leads to ovarian development

To determine whether *figlaY* is necessary for triggering testicular development, we knocked out this gene in XY genotypes using the CRISPR-Cas9 system (fig. S9). Four mosaic G_0 fish carrying mutant alleles were identified, all of which developed as phenotypic males. We then crossed each G_0 males with XX females to produce the G_1 generation. We identified eight different mutant alleles from 87 XY individuals: six frameshift mutations and two nonframeshift mutations (Fig. 3A). All individuals carrying frameshift mutations ($n = 17$) developed as females with normal ovaries (Fig. 3, B and C), whereas all individuals carrying nonframeshift mutations ($n = 29$) developed as males with normal testes (Fig. 3, D and E). Notably, the two nonframeshift mutant alleles, Hap7 and Hap8, which carry deletions of one and three amino acid residues, respectively, within the basic DNA-binding domain, altered the structure of this domain without affecting the protein's function as a male determining factor (fig. S10). This suggests that the basic DNA binding domain of FIGLAY is dispensable, whereas the helix-loop-helix (HLH) domain is essential for its function as a candidate MSD gene. This loss-of-function analysis demonstrates that *figlaY* is essential for male development in Mozambique tilapia with the LG1 sex-determination system.

figlaY evolution is driven by TEs

We analyzed the ~26-kb insertion sequence in the Y chromosome by aligning it to the FishTEDB TE database and annotated an unclassified long terminal repeat (LTR)–type retrotransposon along with two inversely oriented non-LTR retrotransposons, both homologous to RNA-directed DNA polymerase from jockey-like mobile elements, flanking *figlaY* (fig. S11). This arrangement resembles the "jumping" sex-determining locus in salmonid fishes and *Takifugu* pufferfish, which has transposed or translocated to different ancestral autosomes through the action of TEs (*11*, *30*, *31*), suggesting that the *figlaY*-containing sex-determining locus may have also been mobilized by TEs. Notably, more than 62% of the insertion sequence is derived from TEs, whereas most of the remaining regions consist of conserved noncoding repetitive elements, making the entire sequence highly repetitive. TEs are known to disrupt genes (*32*, *33*), drive the emergence of novel genes through transposon capture or exon fusion (*34*), and be exapted as regulatory units, introducing epigenetic modifications and functional elements such as enhancers, repressors, and promoters (*13*–*15*, *35*). On the basis of these observations, we hypothesized that TEs played a critical role in the evolution of *figlaY* into an MSD gene. Consistent with this hypothesis, we found that *figlaY* captured and incorporated highly repetitive TE-derived sequences from across the genome into its structure including its coding region, stop codon, and entire 3′ untranslated region (3′UTR) (Fig. 4A). Further analysis of five conserved noncoding elements (CNEs) and predicted promoter regions within this insertion (Fig. 4, B and C) revealed that the promoter (P6), located directly upstream of *figlaY*, presented substantial transcriptional activity in luciferase reporter assays in an Asian seabass epithelial-like cell line (Fig. 4D). This putative promoter consists of ~45% (141 bp) sequence derived from an unannotated TE family, whereas the remaining 171 bp is unique in the genome. Both elements exhibit significant, albeit reduced, promoter activity compared to the full-length promoter sequence (fig. S12). Notably, this sequence bears no homology with the upstream regulatory region of the autosomal *figla* gene, suggesting that *figlaY* has evolved a distinct promoter through TE-mediated sequence recruitment. Additionally, two of the five CNEs, CNE3 and CNE5, displayed enhancer activity (Fig. 4E). These CNEs were surrounded by TE sequences derived from the LINE/L2 retrotransposon superfamily. Further analysis of three additional upstream TEs

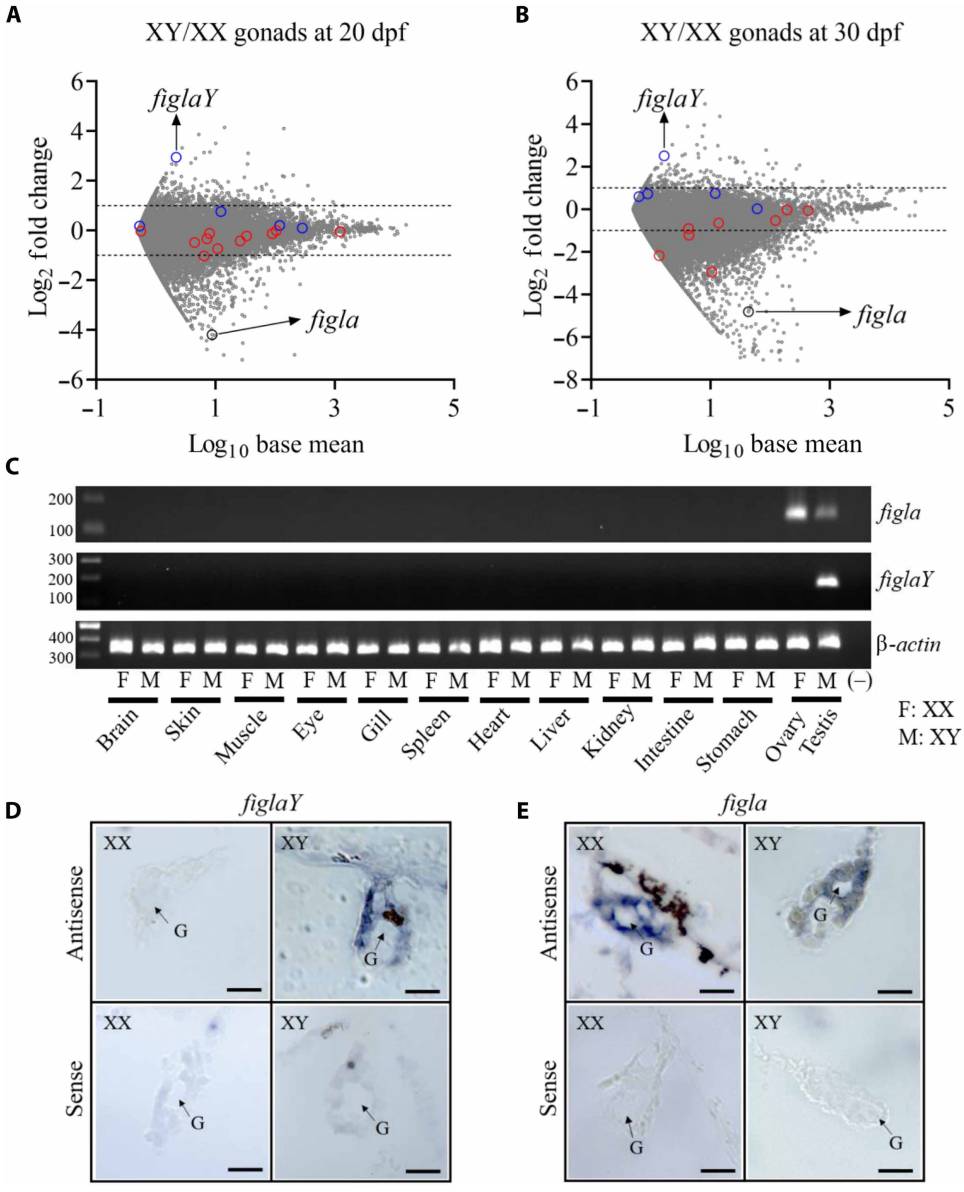

Fig. 2. *figlaY* exhibits male-biased expression during critical stages of sex determination. (**A**) Differentially expressed genes in pooled gonad samples between XY and XX genotypes (one pool from 48 gonads per genotype) at 20 days postfertilization (dpf). Only *figlaY* shows male-biased expression with $\log_2$ (fold change) > 1, whereas its autosomal homolog *figla* exhibits female-biased expression. The threshold for differential expression ($\log_2$ fold change) = 1 is indicated by dotted lines. Genes with higher expression in XX and XY samples are highlighted by red and blue dots, respectively. (**B**) Differentially expressed genes in pooled gonad samples between XY and XX genotypes (one pool from 48 gonads per genotype) at 30 dpf. *figlaY* remains the only gene with male-biased expression, whereas *figla* continues to show female-biased expression. Three additional genes above the threshold line ($\log_2$ fold change = 1) show female-biased expression. (**C**) Expression of *figlaY* and *figla* in various tissues of adult females (XX) and males (XY), detected by reverse transcription PCR (RT-PCR) using gene-specific primers. The housekeeping gene β-*actin* serves as a positive control. An equal amount of RNA from each of three individuals per genotype was pooled to create each sample. (**D**) and (**E**) In situ hybridization (ISH) analysis of *figlaY* and *figla* expression, respectively, in XX and XY gonads at 20 dpf. *figlaY* is detected exclusively in somatic cells surrounding germ cells in XY gonads, whereas *figla* is expressed in somatic cells surrounding germ cells in both XX and XY gonads but shows a weaker signal in XY gonads. G, germ cell; scale bars, 10 μm.

revealed that two, TE49 and TE51, originating from the LINE/L2 and DNA/hAT superfamilies, respectively, also exhibited enhancer activity (Fig. 4F).

Overall, these findings suggest that TEs played a critical role in the emergence and functional diversification of *figlaY* as an MSD gene in tilapia. TEs appear to have facilitated the duplication and translocation of the ancestral *figla* gene, providing the substrate for subsequent neofunctionalization (*36*). In addition, TE insertions likely influenced the structure of *figlaY*, including the introduction of a premature stop codon, a feature commonly associated with the functional divergence of duplicated genes (*37*). TE-derived regulatory sequences may also have contributed to shaping the temporal and tissue-specific expression of *figlaY*, a characteristic shared by several other MSD genes (*15*). Similar roles of TEs in modifying the

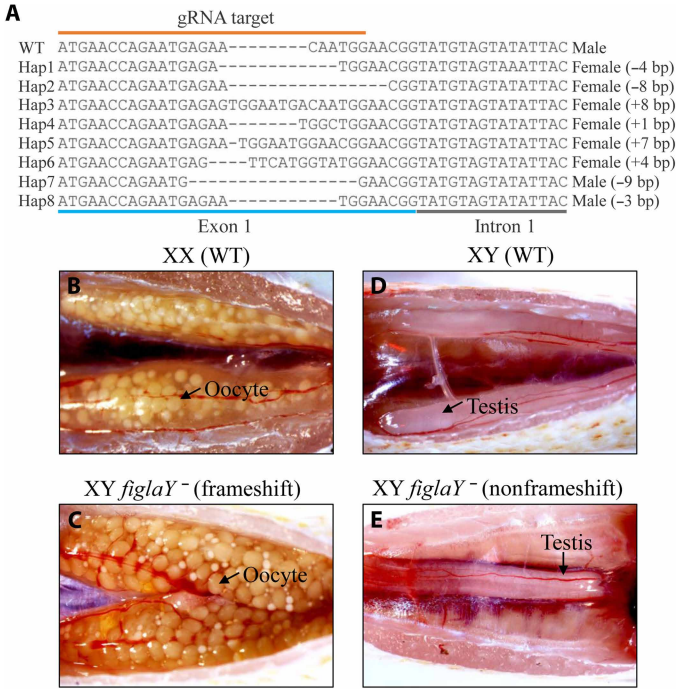

Fig. 3. CRISPR-Cas9 knockout of *figlaY* induces female development in XY tilapia. (**A**) Eight mutant alleles were identified in G_1 CRISPants with XY genotypes, including six frameshift and two nonframeshift variants. The locations of the designed gRNA, exon 1, and intron 1 are indicated. WT, wild type. (**B**) The gonad of an XX wild-type tilapia develops into an ovary at 90 dpf. (**C**) The gonad of an XY CRISPant carrying a frameshift allele develops into an ovary at 90 dpf. (**D**) The gonad of an XY wild-type tilapia develops into a testis at 90 dpf. (**E**) The gonad of an XY CRISPant carrying a nonframeshift allele develops into a testis at 90 dpf. Oocyte and testis are indicated with arrows.

regulation of sex-determining genes such as *dmrt1* and *gsdf* have been reported in sablefish, fighting fish, and medaka, where TE insertions alter gene expression through the introduction or co-option of regulatory elements (*15*). Given that TE integration sites are largely random across the genome (*38*), the evolution of *figlaY* into an MSD gene likely required a specific combination of several mutational events, which may explain why similar complex outcomes have not been repeatedly observed in other vertebrates.

FIGLAY antagonizes the effects of FIGLA in ovarian development

As a truncated homolog, *figlaY* evolved a novel exon 1 while retaining exon 2 and part of exon 3, both of which are highly similar in sequence to the corresponding exons of *figla* (Fig. 5A). Protein sequence analysis revealed that FIGLAY preserves the HLH domain, which mediates protein heterodimerization with other HLH proteins (*39*), but has lost the basic DNA binding domain contained in FIGLA (Fig. 5B and figs. S13 and S14). In human and mice, Figla heterodimerizes with transcription factor E12 and binds to E-box motifs in zona pellucida genes 1-3 (Zp1-3) to promote folliculogenesis (*39*, *40*). In both zebrafish and Nile tilapia with LG23 sex-determination system, FIGLA is essential for ovarian differentiation by antagonizing testis differentiation (*41*). Notably, *figlaY* expression overlaps with *figla* in both time and space during gonadal development in XY genotypes of the Mozambique tilapia studied here, from undifferentiated to differentiated stages. This observation led us to hypothesize that FIGLAY functions as an MSD gene through a "dominant negative" mechanism, similar to the W-linked sex-determining gene *Dm-W* in *Xenopus laevis*, which determines female sex by antagonizing the male-promoting gene *dmrt1* and thereby directing gonadal development toward ovaries (*42*). In the present case, FIGLAY is likely to compete with FIGLA for HLH protein binding, forming heterodimers with reduced DNA-binding capacity. This mechanism may inhibit ovarian differentiation, ultimately leading to testicular development (Fig. 5C).

To test this hypothesis, we cloned the predicted promoter region (~1.3 kb) of *zp2* containing 12 E-box motifs (CANNTG) into a luciferase reporter system. We found that Figla alone could not activate the *zp2* promoter activity, whereas E12 alone did. Cotransfection of FIGLA and E12 further enhanced *zp2* promoter activity compared to E12 alone, in agreement with previous observations in mouse (Fig. 5D) (*39*) and suggesting that the ZP2/E12 heterodimer function is conserved across vertebrates. Notably, FIGLAY significantly suppressed *zp2* promoter activity induced by FIGLA and E12 (Fig. 5E) but had no effect on E12 alone (Fig. 5F). These findings support the hypothesis that FIGLAY antagonizes FIGLA's function in regulating ovarian differentiation.

Functional antagonism between paralogs arising from gene duplication is a common phenomenon in regulatory networks, signaling pathways, and developmental processes (*43*, *44*). Our findings reveal a case in which a duplicate of a femaleness-associated gene evolved into a male MSD gene through antagonistic interaction, representing a previously unrecognized mechanism in vertebrate sex determination. Although ZP proteins have been implicated in folliculogenesis, their precise role in ovarian differentiation remains unclear (*45*, *46*). While we cannot rule out the possibility that Figla heterodimerizes with HLH proteins other than E12 and activates downstream targets beyond ZP proteins in ovarian differentiation, the proposed antagonistic mechanism remains relevant and provides a compelling framework for future investigation. Further studies are therefore needed to elucidate the signaling pathway through which Figla regulates ovarian differentiation and to validate this hypothesis.

figlaY evolved into an MSD gene in an ancient lineage and introgressed into diverse cichlid lineages

Tracing the origin of MSD genes in vertebrates is crucial for understanding the diversity, plasticity, and evolutionary turnover of sex-determination systems (*1*, *16*). In our search of all available cichlid genome assemblies and published literature, we identified, in addition to the *figla* gene, a *figla-like homolog* present in specific lineages, including 11 tilapia species, 39 species from Lake Tanganyikan cichlid tribes, and the relatively ancient African cichlid *Etia nguti* (fig. S15). However, all these genome assemblies containing *figla-like homolog* were highly fragmented (table S6). Additionally, exon 1 of *figla-like homolog* appears to reside in a highly repetitive genomic region, as seen of *figlaY* in Mozambique tilapia, making it unassembled in these species. Moreover, exon 3 of *figla-like homolog* was fragmented or incomplete in three species, which were therefore excluded from further analysis (fig. S15).

Phylogenetic analysis based on the protein sequences corresponding to exons 2 and 3 revealed that the FIGLA-like homolog is an ortholog of the Mozambique tilapia FIGLAY. The analysis also identified two major clades of FIGLA-like homologs: one comprising tilapia lineages and the other representing cichlid tribes from Lake

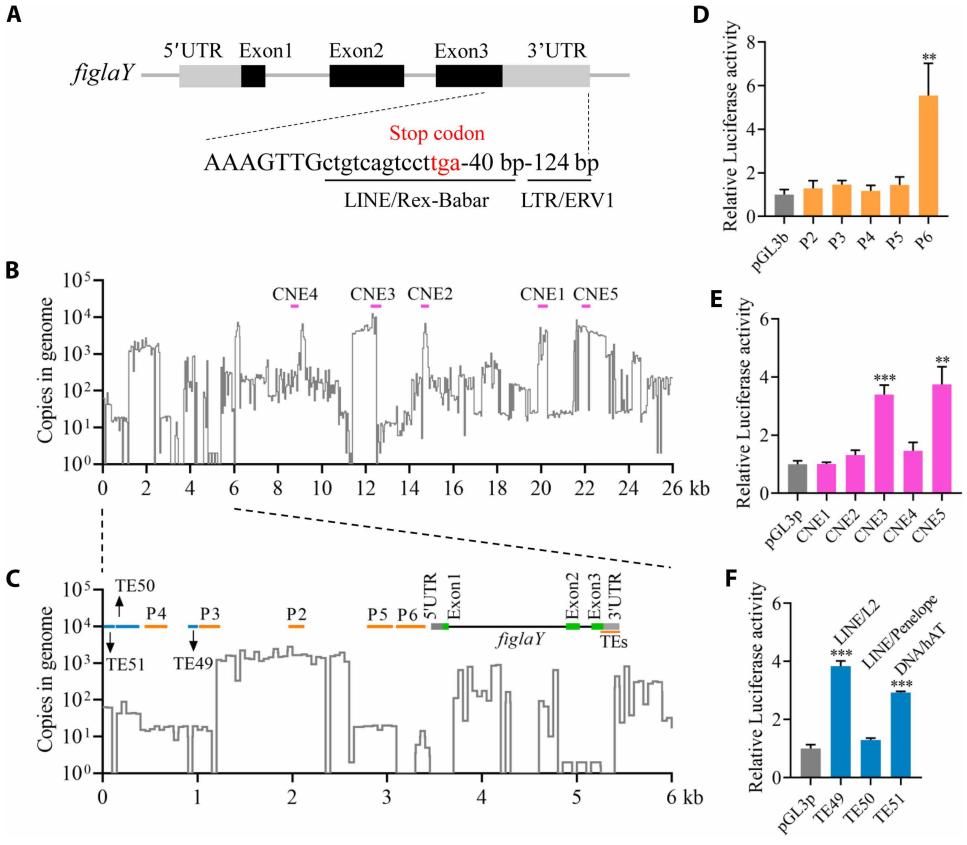

Fig. 4. TEs drive the evolution of *figlaY* into a master sex-determining gene. (**A**) *figlaY* integrates TEs from the LINE and LTR superfamilies into its coding sequences in exon 3 and the 3′ untranslated region (3′UTR). (**B**) The schematic diagram shows the positions of five predicted conserved noncoding elements (CNEs) across teleost species within the *figlaY* locus (pink). The *y* axis indicates copy numbers of repetitive genomic elements, estimated using 50-bp nonoverlapping windows (left *y* axis). (**C**) The schematic diagram illustrates the *figlaY* gene structure, highlighting the positions of five predicted promoter elements (orange) and three TEs selected for functional validation (blue). *y* axis as in (B). (**D**) The predicted promoter element closest to the 5′UTR exhibits significant promoter activity in reporter gene assays. (**E**) Two of the five predicted CNEs show significant enhancer activity in reporter gene assays. (**F**) Two of the three TEs, originating from the LINE/L2 and DNA/hAT superfamilies, exhibit significant enhancer activity in reporter gene assays. (Triplicate transfections were performed for each sample; $^{**}P < 0.01$ and $^{***}P < 0.001$, two-tailed Student's *t* test.)

Tanganyika. The FIGLA-like homolog of *E. nguti*, a relatively ancient lineage, occupied a basal position, suggesting it as the likely ancestral origin of the FIGLA-like group (fig. S16). However, phylogenetic relationships inferred from the highly conserved FIGLA sequences were poorly resolved because of a lack of informative variation in this region across studied species (fig. S16). To improve resolution, we reconstructed the phylogeny using coding sequences of both exons 2 and 3 corresponding to *figlaY*. This approach yielded a clearer and more robust cichlid phylogeny (*47*, *48*), revealing that the *figla-like homolog* in *E. nguti* diverged from the canonical *figla* gene of the same species (Fig. 6). Notably, we identified a *figla-like homolog* in *Coptodon bakossiorum*, which, although incomplete (fig. S15), was phylogenetically positioned between *E. nguti* and the other tilapia and Tanganyikan cichlid species (fig. S17). These findings support the hypothesis that *figlaY* originated through gene duplication in *E. nguti* or a closely related species belonging to a lineage more ancient than tilapia. In medaka (*Oryzias latipes*), gene duplication and translocation occurred simultaneously, without evidence of intermediate introgression (*12*). Similarly, it is likely that the translocation of *figlaY* closely followed its duplication from the autosomal *figla*. However, due to the lack of chromosome-level genome assemblies, particularly for *E. nguti* and *C. bakossiorum*, it remains unclear whether *figlaY* was originally translocated to the chromosome homologous to Mozambique tilapia LG1 or whether it underwent additional translocation events before lastly integrating into that chromosome.

Beyond Mozambique tilapia, the *figla-like homolog* has also been specifically identified in males of seven additional tilapia species or strains that possess an LG1 sex-determination system. These include species from relatively basal and geographically widespread cichlid tribes such as Coptodonini, Oreochromini, Pelmatolapiini, and Heterotilapiini (Fig. 6) (*20*, *49*). Thus, *figla-like homolog* appears to be present in only a limited subset of tilapia lineages. In Lake Tanganyika, we detected *figla-like homolog* in 39 of more than 200 species with available genome sequences (*49*). Of these, 35 species had genome data from both sexes (table S6). Notably, *figla-like homolog* was found exclusively in the male genome sequences of 32 of the 35 species (91.4%) (Fig. 6 and table S6), strongly suggesting a male-specific role. These findings imply that *figlaY* may act as a sex-determining gene in some Tanganyikan species as well.

Overall, the patchy taxonomic distribution, phylogenetic placement, and strong male specificity of *figlaY* across African cichlids indicate that it represents an evolutionarily mobile male-determining factor rather than a neutral lineage marker. These patterns strongly

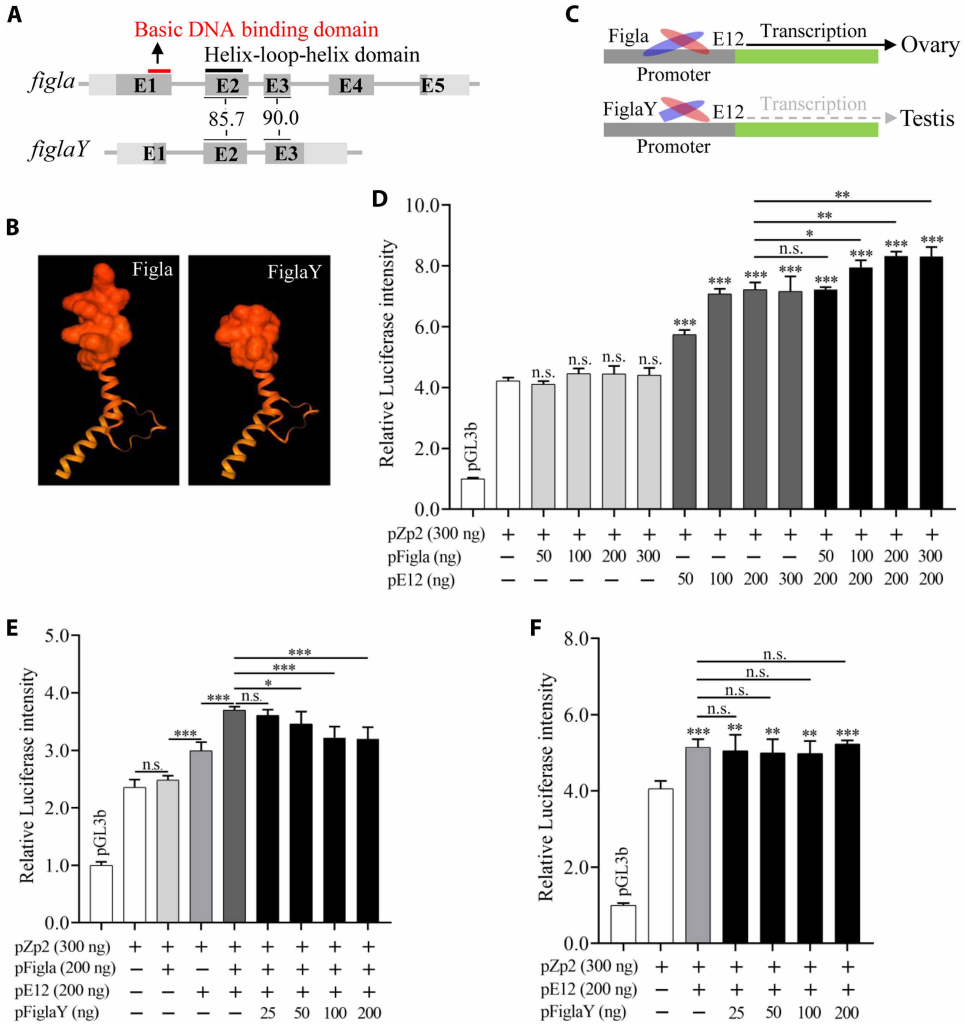

Fig. 5. FIGLAY antagonizes FIGLA-mediated activation of the zp2 promoter by in the presence of E12. (**A**) Sequence similarity comparison between FIGLA and FIGLAY, showing homology (%) restricted to exon 2 and exon 3. The basic DNA binding domain and HLH domain of FIGLA are indicated. (**B**) Predicted 3D structures of exon 1 of FIGLA and FIGLAY, highlighting the absence of the basic DNA binding domain in FIGLAY. (**C**) Proposed model of FIGLAY function. FIGLA forms a heterodimer with E12, enabling binding to and activation of the zp2 promoter. In contrast, FIGLAY can also bind to E12, but the resulting heterodimer lacks the basic DNA binding domain and therefore cannot interact with the zp2 promoter, preventing activation of its expression. (**D**) Luciferase assays showing that FIGLA and E12 synergistically enhance zp2 promoter activity. FIGLA alone does not induce activation, while E12 alone elicits moderate activation. Assays were performed with varying concentrations of FIGLA plasmid (50 to 300 ng), E12 plasmid (50 to 300 ng), or 200 ng E12 combined with increasing FIGLA plasmid (50 to 300 ng). zp2 plasmid alone served as controls ($n = 3$ independent transfections). (**E**) FIGLAY inhibits activation of the zp2 promoter by the FIGLA/E12 complex. Luciferase assays were performed with 200 ng each of E12 and FIGLA plasmids plus increasing FIGLAY plasmid (0 to 200 ng). Samples without FIGLAY plasmid served as controls ($n = 6$ independent transfections). (**F**) FIGLAY does not interfere with zp2 promoter activation by E12 alone. Luciferase assays were performed with 200 ng of E12 plasmid plus increasing FIGLAY plasmid (0 to 200 ng). Samples without FIGLAY plasmid served as controls ($n = 4$ independent transfections). Statistical significance was determined using an unpaired two-tailed Student's *t* test. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; n.s., not significant.

support a model in which *figlaY* originated once via duplication and translocation in an ancient lineage, followed by lineage sorting and extensive introgression. In contrast, ancestral polymorphism and horizontal gene transfer make sharply different predictions regarding gene age, sex specificity, and phylogenetic structure that are inconsistent with the observed data (*50*–*52*). The presence of *figlaY* in the phylogenetically basal cichlid *E. nguti* but only in subsets of the Tilapia and Lake Tanganyika lineages indicates an ancient origin coupled with differential retention. Its near-exclusive occurrence in male genomes further rules out ancestral polymorphism, as long-persisting ancestral alleles would be expected to segregate in both sexes (*51*). The restricted yet phylogenetically dispersed distribution of *figlaY* is most parsimoniously explained by introgression. In particular, genomic evidence suggests that West-Central African riverine ancestors of tilapia lineage colonized Lake Tanganyika and hybridized with early resident tribes at the base of the radiation (*48*), potentially facilitating the introgression of *figlaY*. As a dominant male-determining gene, *figlaY* would have a transmission advantage, enabling the rapid replacement of resident sex-determination systems across species. Under this model, the *figla*-like homolog evolved into the male sex-determining gene *figlaY* in an ancestor lineage of *E. nguti* or a closely related riverine species, behaving as a selfish, adaptive Y-linked

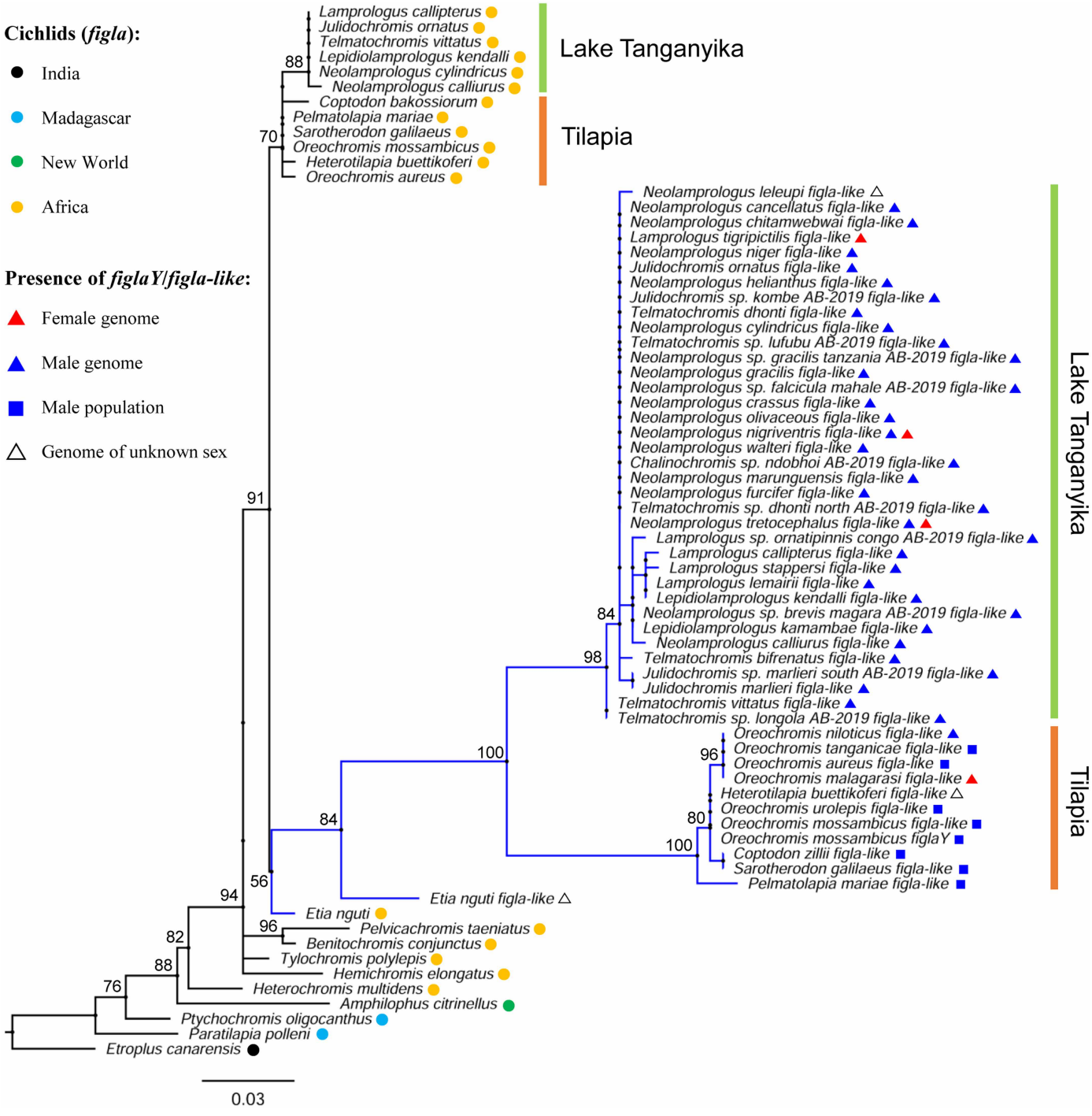

Fig. 6. *figlaY* originated from a gene duplication event in a relatively ancient cichlid species. A phylogenetic tree was constructed using the coding sequences of *figla* and *figla-like homologs*, including Mozambique tilapia *figlaY*, corresponding to the exons 2 and 3 region of the *figlaY* gene. Species origins are indicated by solid colored dots. The presence of *figlaY* or *figla-like homologs* in male- or female-specific genome assemblies or populations of each species is indicated by colored triangles or squares. Branches representing *figla-like homologs*, including Mozambique *figlaY* and its closest ancestral *figla*, are highlighted in blue. Bootstrap values from 1000 replicates are shown at the corresponding branches. Scale of sequence evolutionary rate is indicated. The scale bar represents substitutions per site. Branch lengths are proportional to the rate of sequence evolution.

element capable of crossing species boundaries. Moreover, the coherent, lineage-resolved phylogeny and sex-linked genomic context of *figlaY* are incompatible with horizontal gene transfer, which, in vertebrates, is limited to short, unlinked DNA fragments and produces mosaic rather than syntenic, chromosome-based inheritance patterns (*52*, *53*).

Together, these results support a scenario in which *figlaY* arose via duplication and translocation in a relatively ancient cichlid species, likely in a common ancestor with the *E. nguti* lineage. It subsequently evolved into an MSD gene with the assistance of TEs and was later introgressed into several cichlid lineages, including tilapia and certain Lake Tanganyika species. This case provides a compelling example of how gene duplication and introgression can drive the emergence and spread of novel sex-determining mechanisms in vertebrates. More broadly, introgression can facilitate the rapid spread of new MSD genes across species, promoting sex chromosome turnover, diversification of sex-determination systems, and reproductive isolation. This process might be common throughout African cichlid radiations (*54*, *55*). The *figlaY* system thus illustrates how introgression acts as a powerful force of convergence in sex determination, providing a pathway for adaptive solutions to cross species boundaries and shape hybridization outcomes.

In the case of MSD genes arising from gene duplication or other large structural variations, the lack of sequence homology between proto-sex chromosomes at the MSD gene region would result in the accumulation of sequence divergence over time (*37*, *56*, *57*). This divergence allows us to estimate the relative timing of *figlaY* translocation. We analyzed the X and Y chromosome sequences and identified a highly differentiated region (HDR) of ~1.3 Mb upstream of the SDR on the Y chromosome, while no substantial differentiation was observed in the SDR except for the 26-kb insertion (Fig. 7 and fig. S18). The SDR on the Y chromosome was inverted and translocated from

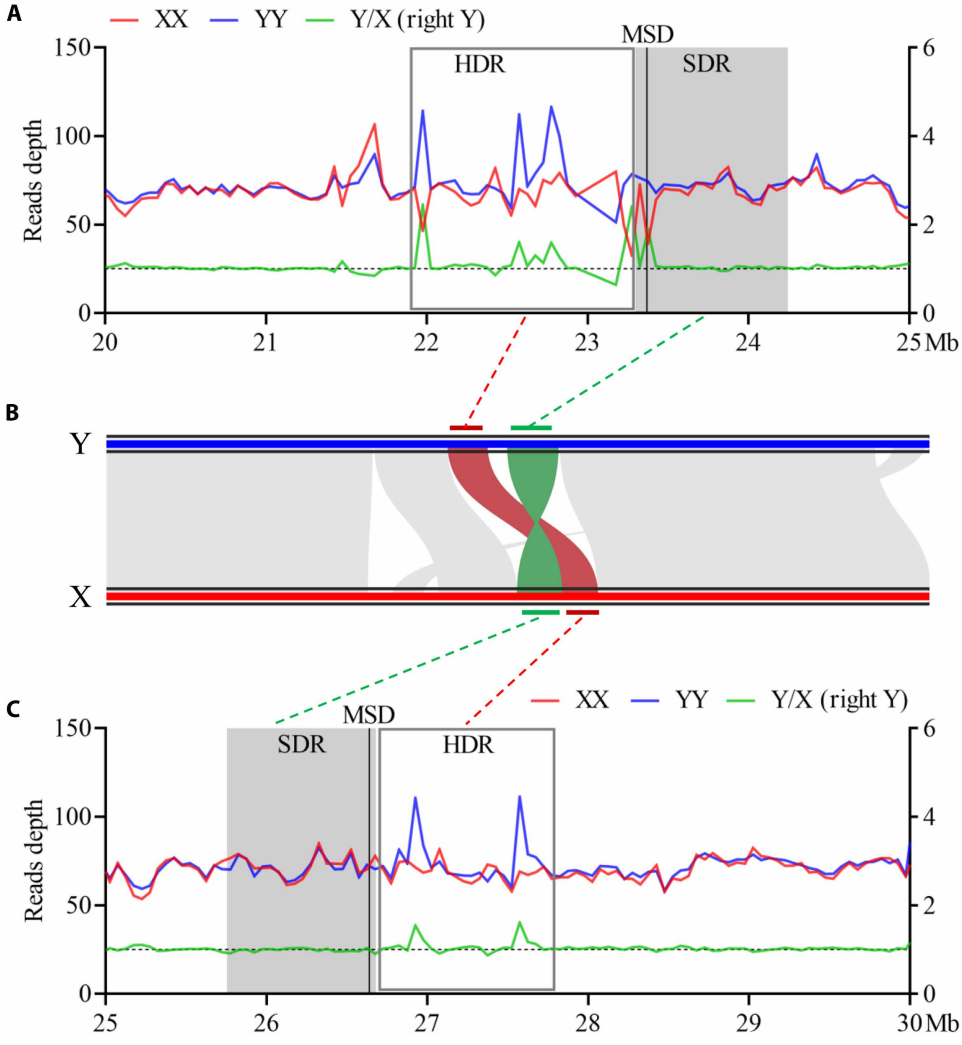

Fig. 7. Elevated sequence divergence in the flanking region but not in the SDR between X and Y chromosomes. (**A**) Sequencing read coverage from XX and YY genotypes mapped to the Y chromosome, estimated using 50-kb nonoverlapping windows (left *y* axis). The ratio of Y-specific to X-specific read coverage is shown (right *y* axis). The SDR, MSD gene, and HDR are indicated by a shaded area, a vertical line, and a box, respectively. (**B**) An inverted relocation event is identified between the X and Y chromosomes. The translocated region overlaps with both the SDR and HDR, with the inversion specifically coinciding with the SDR. (**C**) Sequencing read coverage from XX and YY genotypes mapped to the X chromosome, estimated using 50-kb nonoverlapping windows (left *y* axis). The ratio of Y-specific to X-specific read coverage is shown (right *y* axis). The SDR, MSD, and HDR are highlighted as in (A).

the upstream to the downstream of the HDR compared to its location on the X chromosome (Fig. 7B). This inverted relocation is known to disrupt sequence homology between sex chromosomes, suppress recombination, and drive genetic differentiation, which are processes commonly associated with sex chromosome divergence and degeneration (*12*, *14*). Consistent with this, we observed extended recombination suppression on both sex chromosomes, particularly the Y chromosome (fig. S19). Sequence annotation revealed that the HDR on the Y chromosome contained a higher proportion of TEs than the corresponding region on the X chromosome (32.3 versus 28.7%) (fig. S20), suggesting that TE expansion in Y chromosome also contributed to sequence divergence in this region. In contrast, the TE content in the SDR was only slightly higher on the Y chromosome compared to the X chromosome (20.7 versus 19.0%) (fig. S20). This suggests that TE amplifications may have occurred simultaneously in the SDR of the Y chromosome and the corresponding region of the X chromosome, as reported in some other teleost species (*58*). Additionally, differences in methylation profiles or chromatin conformation resulting from TE expansions between the sex chromosomes may also contribute to recombination suppression in this region (*21*). Furthermore, the synonymous substitution rate (K_s) values of genes in the SDR showed no significant difference from those in the flanking HDR or the remaining regions (fig. S21), indicating that sequence divergence primarily occurred in the HDR, while divergence in the SDR remains minimal, compatible with an early stage of differentiation. This also suggests a recent turnover of the sex chromosome, as observed in some species, where MSD genes have emerged from recent sequence structural variations, accompanied incipient sex chromosome evolution (*14*, *59*). Together, our findings provide a detailed account of the evolutionary trajectory leading to the emergence of *figlaY* as an MSD gene in cichlids. Given the patterns of chromosome rearrangement and sequence differentiation, we cannot rule out the possibility that the MSD gene has recently translocated from the HDR to the SDR, likely due to TE activities.

MATERIALS AND METHODS

Genome sequencing of XX and YY fish

The Mozambique tilapia (*O. mossambicus*) used in this study were introduced into Singapore for selective breeding in 2010 from wild populations in the east-flowing river systems of South Africa that drain into the Indian Ocean. Previous study identified the presence of the MSD gene on LG1 in this population (*26*). To establish a breeding cross, we first genotyped the broodstock using two microsatellite markers (Omo1314 and S154) linked to phenotypic sex (*26*). We selected one XX female and one XY male for mating. After hatching, ~200 fry were raised in a 60-liter tank with recirculating fresh water in our fish facility. They were fed three times daily with a standard commercial fry diet (Nissin, Japan), supplemented with 17β-estradiol (60 mg/kg; Sigma-Aldrich, Germany). After 1 month, the F1 fry were transferred to a 1000-liter tank and fed with untreated feed. Once sexually mature, each fish's phenotypic and genotypic sex was determined. A sex-reversed XY female was selected and crossed with an XY male to produce F2 YY fish. Upon reaching sexual maturity, we observed that all YY individuals developed as males, except for a few with abnormally developed genital pores.

One YY male from the F2 family and one XX female from a normal family were selected for genome sequencing using the PacBio Sequel II (Pacific Biosciences, USA) and Nanopore PromethION (Oxford Nanopore Technologies, UK) platforms, respectively. Sequencing was performed by NovogeneAIT Genomics, Singapore (table S7). Genomic DNA for long-read sequencing was extracted from fin tissues using the standard phenol-chloroform method. Additionally, both samples, along with an XY male, were subjected to whole-genome sequencing on the Illumina NovaSeq 6000 platform (Illumina, USA) (table S7). To generate chromosome-level genome assemblies, Hi-C libraries were constructed for the XX and YY fish using blood samples, following a modified version of our previously published protocol (*14*). These libraries were sequenced on the Illumina NovaSeq 6000 platform for 150-bp paired-end reads (table S7).

Genome assembly and annotation

Both PacBio and Nanopore reads were cleaned and assembled into contigs using Flye v2.8.2b (*60*) with the following parameters: --pacbio-raw/--nano-raw --asm-coverage 50 -g 1 g -i 2. To reduce assembly redundancy, Purge_Haplotigs (*61*) was applied with default parameters. The assembled contigs were then polished with Pilon v1.24 (*62*) using corresponding Illumina short reads. Hi-C reads were used to anchor contigs to scaffolds using the Juicer v1.6 (*63*) and 3D-DNA v180922 (*64*) pipelines. The scaffolds were then manually curated using Juicebox v2.16 (*63*), with a prior setting of 22 haploid chromosomes for this species (*26*). Chromosome names were assigned following the Nile tilapia nomenclature. Genome assembly completeness was assessed using BUSCO v5.2.2 (*65*) against the actinopterygii_odb10 database.

Protein coding genes were predicted using the Maker2 pipeline (*66*). First, repetitive elements were soft masked using RepeatMasker v4.0.7 (*67*) based on the Repbase database (*68*), a custom library built with RepeatModeler v2.0 (*69*), and short tandem repeats identified by Tandem Repeat Finder v4.09 (*70*). Gene predictions were based on protein homology and Expressed sequence tag (EST) evidence from closely related species. Protein and transcript sequences were retrieved from GenBank for Nile tilapia (*O. niloticus*, GCA_001858045.3), Blue tilapia (*Oreochromis aureus*, GCA_013358895.1), *Maylandia zebra* (GCA_000238955.5), *Haplochromis nyererei* (GCA_000239375.1), and *Haplochromis burtoni* (GCA_018398535.1). Additionally, we sequenced the transcriptomes of one ovary and one testis from XX and XY fish, respectively. Transcripts were assembled de novo using Trinity v2.15 (*71*) and incorporated as EST evidence. Gene models were iteratively trained over three rounds using SNAP v2013 (*72*) and Augustus v3.2.3 (*73*).

Mapping the sex-determining locus

A mapping family was generated by crossing an XX female with an XY male. Offspring were raised under normal laboratory conditions until sexual maturity. A total of 288 individuals were randomly selected for sex phenotyping via histological analysis, revealing a 1:1 female-to-male ratio (147 females and 141 males). Fin tissue was collected for DNA isolation using the DNeasy Blood & Tissue Kit (QIAGEN, Germany). The entire mapping family was genotyped using Restriction site-Associated DNA sequencing (RADseq) technology (*74*), following our previous study (*27*). RADseq libraries were sequenced on an Illumina NextSeq 500 platform, generating 1 × 150-bp reads (table S7). Raw sequencing reads were processed using process_radtags from the Stacks v2.45 package (*75*) with the following parameters: -r -c -q -t 136. Cleaned reads were aligned to the YY genome using BWA-MEM v0.7.17 (*76*) with default settings, retaining only uniquely aligned reads for further analysis. SNP discovery and genotyping were performed using Stacks v2.45 package,

keeping only one SNP per RADtag (*75*). After filtering, 244 samples with high sequence coverage and SNPs present in >200 individuals were retained for further analyses.

SNPs were further filtered for Mendelian segregation distortion using chi-square tests (significance threshold, 0.01), yielding 6850 SNPs. A sex-averaged linkage map was constructed using Lep-MAP3 (*77*) with a Logarithm of the Odds (LOD) cutoff of 10, resulting in 22 LGs. Mapping of the sex-determining locus was performed using MapQTL6 (Kyazma, Wageningen, The Netherlands) under the multiple QTL model mapping algorithm. The LOD threshold for significant QTL was determined by 1000 permutation tests. To further refine the sex-determining locus, a larger population (960 individuals) derived from the same broodstock as the mapping family was genotyped using SNPs and InDels located in the QTL region, which were fixed between females and males (table S8). Genotyping was performed using Sanger sequencing and PCR assays by running agarose gels or capillary electrophoresis. These genotypes were used to identify recombinants at the sex-determining locus.

Development of a sex marker for rapid PCR assays

To rapidly determine genotypic sex, we developed an InDel marker (Tila.sex.47k) for PCR-based assays. Briefly, whole-genome resequencing datasets from XX, XY, and YY genotypes were aligned to the YY genome using BWA-MEM v0.7.17 (*76*). Genetic variants were identified using the GATK v4.2.2.0 pipeline (*78*). InDels located within the nonrecombining SDR and of suitable length were selected for primer design. Sex diagnostic InDels were validated in broodstock using PCR assays, followed by 2.0% agarose gel electrophoresis. Some of these diagnostic InDels were also used for fine mapping the sex-determining locus, as described above.

Gene expression analysis by transcriptome sequencing

Gonadal sex differentiation between genotypes ($n = 3$) was examined histologically using serially cross-sectioned trunks (7 μm) stained with hematoxylin and eosin at different developmental stages, following our previous study (*14*). To investigate gene expression patterns, both undifferentiated and differentiating gonads were dissected separately from 48× XX and 48× XY genotypes. For these fish, heads or tails were used for DNA extraction and genotyping. Gonad samples were pooled into four samples for RNA extraction using TRIzol reagent (Invitrogen, USA). These RNA samples were then used to construct mRNA sequencing libraries with the Illumina TruSeq RNA Library Prep Kit v2 (Illumina, USA). The mRNA libraries were sequenced on an Illumina NextSeq 500 platform, generating either 2× 150-bp or 1× 150-bp reads (table S7). Additionally, transcriptomes from one ovary and one testis of XX and XY adults, respectively, were sequenced using the same method for genome annotation. Raw sequencing reads were processed using process_shortreads from the Stacks v2.45 package (*75*) with the parameters: -r -c -q -t 150. Cleaned reads were then mapped to the YY genome using STAR v2.5.2b (*79*) with default settings. Gene transcripts in each sample were quantified using HTSeq-count v0.9.1 (*80*), and transcript counts across samples were normalized by transcripts per kilobase million.

Gene expression analysis by RT-PCR

Gene expression patterns were analyzed using RT-PCR. Total RNA was extracted from tissues or embryos of the same genotype, as described above. Two micrograms of RNA were treated with deoxyribonuclease I (Roche, Switzerland) and reverse transcribed into cDNA using M-MLV Reverse Transcriptase (Promega, USA). Relative gene expression was assessed using RT-PCR and qRT-PCR with gene specific primers (table S8), following our previous study (*14*). For embryo/truck samples, cDNA synthesized from 500 ng of total RNA (pooled from 12 individuals per genotype) was used. For independent tissue samples, cDNA synthesized from 20 ng of total RNA was used. β-*Actin* was used as the housekeeping gene for normalization. qRT-PCR was performed using KAPA SYBR FAST qPCR Kits (KapaBiosystems, USA) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, USA) under standard PCR conditions: 35 cycles with an annealing temperature of 60°C. Each sample was analyzed in three biological replicates, with three technical replicates per biological sample. Relative gene expression levels were calculated using the $2^{-\Delta\Delta CT}$ method.

Gene expression analysis by In-situ hybridization (ISH)

Embryos of different genotypes at early development stages were fixed in 4% paraformaldehyde. For ISH, cDNA fragments spanning the partial 5′UTR and first exon of *figlaY*, as well as the partial 3′UTR of *figla*, were amplified separately and cloned into the pGEM T-easy vector (Promega, USA). Antisense and sense digoxigenin (DIG)–labeled RNA probes were synthesized via in vitro transcription using the DIG RNA Labeling Kit (Roche, Switzerland). ISH was performed following a previous method (*81*) with some modifications. In brief, trunk sections were deparaffinized and hydrated. Sections were digested with proteinase K (2 mg/ml; Sigma-Aldrich, Germany) at 37°C for 20 min. Hybridization was carried out with DIG-labeled RNA probes at 57°C for more than 20 hours. Hybridization signals were detected using 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium via the DIG Nucleic Acid Detection Kit (Roche, Switzerland). Visualization was performed using a Zeiss Axioplan 2 microscope system.

CRISPR-Cas9 knockout of *figlaY*

A guide RNA (gRNA) targeting the first exon of *figlaY* was designed using E-CRISP (*82*) and synthesized following our previous study (*83*). The gRNA was validated in vitro by digesting genomic DNA with the Cas9 Nuclease NLS enzyme [New England Biolabs (NEB), USA]. For microinjection, Cas9 Nuclease NLS protein (NEB, USA) and gRNA were prepared at final concentrations of 200 ng/μl each and coinjected into one-cell stage embryos produced from crosses between XX female and XY male parents. To screen for genetic modifications, site-specific primers flanking the targeted sequence were designed, and PCR products were sequenced using Sanger sequencing and Thymine-Adenine (TA) cloning. Four mosaic G_0 CRISPants with genetic modifications at and around the target site were obtained and raised to mature (all males). These G_0 males were crossed with XX females to produce G_1 fish. At 3 months postfertilization, G_1 individuals were screened for mutants using Sanger sequencing and TA cloning. G_1 individuals carrying both frameshift and nonframeshift mutations were euthanized and dissected for histological analysis of gonadal development, along with wild-type individuals of different genotypes.

Luciferase reporter assay

Putative promoter and enhancer sequences were predicted using the program Promoter v2.0 (*84*) and by mapping to the tilapia CNEs database (*85*). These sequences were cloned into the pGL3-basic and pGL3-promoter vectors (Promega, USA), respectively. Each construct

(200 ng) was cotransfected with 300 ng of pRL Renilla Luciferase Control Reporter Vector (Promega, USA) into an Asian seabass epithelial-like cell line (*86*) cultured in 24-well plates at 70 to 90% confluence. Transfections were performed using Lipofectamine 3000 (Thermo Fisher Scientific, USA). After 48 hours, luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega, USA) on a Tecan Spark microplate reader (Tecan, Switzerland). Three independent transfections were conducted for each sample.

Protein structure predictions were performed using AlphaFold2 (*87*). The promoter sequences of Zp proteins, putative direct targets of Figla, were predicted on the basis of the CANNTG motif as Figla binding sites (*88*–*90*). A ~1.3-kb predicted promoter sequence upstream of the TATA box of one *zp2* gene, containing 12 CANNTG motifs, was subcloned into the pGL3-basic vector. Transcriptome analysis revealed female biased expression of the *zp2* gene in both undifferentiated and differentiating gonads between XX and XY genotypes, with fold changes of ~1.5 and 59.7, respectively. The ORFs of *figla*, *figlaY*, and *e12* were separately cloned into the pcDNA3.1 vector (Invitrogen, USA). Plasmids were extracted using the PureLink HiPure Plasmid Midiprep Kit (Invitrogen, USA) and verified by Sanger sequencing. Human embryonic kidney 293T cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum in 24-well plates at 37°C with 5% CO_2. Transfections were performed at 50 to 70% confluence using Lipofectamine 3000 (Thermo Fisher Scientific, USA). For each well, 1000 ng of plasmid DNA were cotransfected, including 300 ng of pGL3-basic vector containing the *zp2* promoter, 50 to 600 ng of pcDNA3.1 with *figla*, *figlaY*, and/or *e12* ORFs, 0 to 600 ng of empty pcDNA3.1 vector, and 20 ng of pRL-TK (Promega, USA) as a control reporter. Luciferase activity was measured after 48 hours as described above.

Evolutionary analyses

To study the suppression of recombination, sex-specific genetic maps were constructed using the mapping family described above. The genetic positions of markers were plotted against their physical positions to visualize sex-specific differences in recombination (*91*). Sequence divergence between X and Y chromosomes was assessed by aligning whole-genome sequencing reads from XX, XY, and YY genotypes to the X and Y chromosomes, following our previously method (*14*).

For phylogenetic analysis, protein sequences and coding sequence of *figla* homologs from related species were separately aligned using MUSCLE v3.8 (*92*). Multiple sequence alignments were refined using trimAl v1.4 (*93*) to remove gaps and subsequently used to construct a phylogenetic tree using IQ-TREE v2.3.2 (*94*), using the K2P+G4 and HIVb substitution models for protein and coding sequences, respectively, as determined according to Bayesian Information Criterion by ModelFinder (*95*) implemented in IQ-TREE. The K_s for *figla* was calculated between species of interest and Mozambique tilapia, as well as between *figla* and *figlaY* in Mozambique tilapia, using KaKs_calculator v2.0 (*96*) with the Yang-Nielsen method (*97*). Pairwise coding sequence alignments were guided by their corresponding protein alignments using MACSE v2.0 (*98*).

Ethics declarations

All procedures for handling of fish were according to the instructions of the Institutional Animal Care and Use Committee of Temasek Life Sciences Laboratory, Singapore [approval no. TLL(F)-11-001 and TLL(F)-22-004].

Supplementary Materials

This PDF file includes:

Figs. S1 to S21
Tables S1 to S8

REFERENCES

1. D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, Tree of Sex Consortium, Sex determination: Why so many ways of doing it? *PLOS Biol.* **12**, e1001899 (2014).
2. S. Pla, C. Benvenuto, I. Capellini, F. Piferrer, Switches, stability and reversals in the evolutionary history of sexual systems in fish. *Nat. Commun.* **13**, 3029 (2022).
3. Y. Nagahama, T. Chakraborty, B. Paul-Prasanth, K. Ohta, M. Nakamura, Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol. Rev.* **101**, 1237–1308 (2021).
4. Q. Pan, T. Kay, A. Depincé, M. Adolfi, M. Schartl, Y. Guiguen, A. Herpin, Evolution of master sex determiners: TGF-β signalling pathways at regulatory crossroads. *Philos. Trans. R. Soc. B* **376**, 20200091 (2021).
5. J. Kitano, S. Ansai, Y. Takehana, Y. Yamamoto, Diversity and convergence of sex-determination mechanisms in teleost fish. *Annu. Rev. Anim. Biosci.* **12**, 233–259 (2024).
6. M. Matsuda, M. Sakaizumi, Evolution of the sex-determining gene in the teleostean genus *Oryzias*. *Gen. Comp. Endocrinol.* **239**, 80–88 (2016).
7. T. Uller, H. Helanterä, From the origin of sex-determining factors to the evolution of sex-determining systems. *Q. Rev. Biol.* **86**, 163–180 (2011).
8. K. Kikuchi, S. Hamaguchi, Novel sex-determining genes in fish and sex chromosome evolution. *Dev. Dyn.* **242**, 339–353 (2013).
9. A. Herpin, I. Braasch, M. Kraeussling, C. Schmidt, E. C. Thoma, S. Nakamura, M. Tanaka, M. Schartl, Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. *PLOS Genet.* **6**, e1000844 (2010).
10. L. Schrader, J. Schmitz, The impact of transposable elements in adaptive evolution. *Mol. Ecol.* **28**, 1537–1549 (2019).
11. S. Bertho, A. Herpin, M. Schartl, Y. Guiguen, Lessons from an unusual vertebrate sex-determining gene. *Philos. Trans. R. Soc. B* **376**, 20200092 (2021).
12. M. Kondo, I. Nanda, M. Schmid, M. Schartl, Sex determination and sex chromosome evolution: Insights from medaka. *Sex. Dev.* **3**, 88–98 (2009).
13. A. Martin, C. Troadec, A. Boualem, M. Rajab, R. Fernandez, H. Morin, M. Pitrat, C. Dogimont, A. Bendahmane, A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461**, 1135–1138 (2009).
14. L. Wang, F. Sun, Z. Y. Wan, Z. Yang, Y. X. Tay, M. Lee, B. Ye, Y. Wen, Z. Meng, B. Fan, Y. Alfiko, Y. Shen, F. Piferrer, A. Meyer, M. Schartl, G. H. Yue, Transposon-induced epigenetic silencing in the X chromosome as a novel form of *dmrt1* expression regulation during sex determination in the fighting fish. *BMC Biol.* **20**, 5 (2022).
15. A. Herpin, M. Schartl, A. Depincé, Y. Guiguen, J. Bobe, A. Hua-Van, E. S. Hayman, A. Octavera, G. Yoshizaki, K. M. Nichols, G. W. Goetz, J. A. Luckenbach, Allelic diversification after transposable element exaptation promoted *gsdf* as the master sex determining gene of sablefish. *Genome Res.* **31**, 1366–1380 (2021).
16. B. Capel, Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* **18**, 675–689 (2017).
17. A. Herpin, M. Schartl, Plasticity of gene-regulatory networks controlling sex determination: Of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep.* **16**, 1260–1274 (2015).
18. V. Klett, A. Meyer, What, if anything, is a Tilapia?—Mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. *Mol. Biol. Evol.* **19**, 865–883 (2002).
19. J. F. Baroiller, H. D'Cotta, "Sex control in tilapias," in *Sex Control in Aquaculture*, H.-P. Wang, F. Piferrer, S.-L. Chen, Z.-G. Shen, Eds. (Wiley, 2018), pp. 189–234.
20. A. Y. Curzon, A. Shirak, A. Benet-Perlberg, A. Naor, S. I. Low-Tanne, H. Sharkawi, M. Ron, E. Seroussi, Absence of *figla-like* gene is concordant with femaleness in cichlids harboring the LG1 sex-determination system. *Int. J. Mol. Sci.* **23**, 7636 (2022).
21. T. Ge, X. Gui, J.-X. Xu, W. Xia, C.-H. Wang, W. Yang, K. Huang, C. Walsh, J. G. Umen, J. Walter, Y.-R. Du, H. Chen, Z. Shao, G.-L. Xu, DNA cytosine methylation suppresses meiotic recombination at the sex-determining region. *Sci. Adv.* **10**, eadr2345 (2024).
22. B. Lee, G. Hulata, T. Kocher, Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* **92**, 543–549 (2004).
23. A. Cnaani, B.-Y. Lee, N. Zilberman, C. Ozouf-Costaz, G. Hulata, M. Ron, A. D'Hont, J.-F. Baroiller, H. D'Cotta, D. J. Penman, E. Tomasino, J.-P. Coutanceau, E. Pepey, A. Shirak, T. D. Kocher, Genetics of sex determination in tilapiine species. *Sex. Dev.* **2**, 43–54 (2008).
24. A. G. Ciezarek, T. K. Mehta, A. Man, A. G. Ford, G. D. Kavembe, N. Kasozi, B. P. Ngatunga, A. H. Shechonge, R. Tamatamah, D. W. Nyingi, Ancient and recent hybridization in the *Oreochromis* cichlid fishes. *Mol. Biol. Evol.* **41**, msae116 (2024).

25. M. Li, Y. Sun, J. Zhao, H. Shi, S. Zeng, K. Ye, D. Jiang, L. Zhou, L. Sun, W. Tao, Y. Nagahama, T. D. Kocher, D. Wang, A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genet.* **11**, e1005678 (2015).
26. F. Liu, F. Sun, J. Li, J. H. Xia, G. Lin, R. J. Tu, G. H. Yue, A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis spp*.) and mapping of sex-determining loci. *BMC Genomics* **14**, 58 (2013).
27. W. Tao, J. Cao, H. Xiao, X. Zhu, J. Dong, T. D. Kocher, M. Lu, D. Wang, A chromosome-level genome assembly of Mozambique Tilapia (*Oreochromis mossambicus*) reveals the structure of sex determining regions. *Front. Genet.* **12**, 796211 (2021).
28. C. Palaiokostas, M. Bekaert, M. G. Khan, J. B. Taggart, K. Gharbi, B. J. McAndrew, D. J. Penman, Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus L.*) using RAD sequencing. *PLOS ONE* **8**, e68389 (2013).
29. D. Jiang, J. Chen, Z. Fan, D. Tan, J. Zhao, H. Shi, Z. Liu, W. Tao, M. Li, D. Wang, CRISPR/Cas9-induced disruption of wt1a and wt1b reveals their different roles in kidney and gonad development in Nile tilapia. *Dev. Biol.* **428**, 63–73 (2017).
30. J. J. Faber-Hammond, R. B. Phillips, K. H. Brown, Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. *Genome Biol. Evol.* **7**, 1972–1987 (2015).
31. A. Kabir, R. Ieda, S. Hosoya, D. Fujikawa, K. Atsumi, S. Tajima, A. Nozawa, T. Koyama, S. Hirase, O. Nakamura, Repeated translocation of a supergene underlying rapid sex chromosome turnover in *Takifugu pufferfish*. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2121469119 (2022).
32. A. F. Kautt, C. F. Kratochwil, A. Nater, G. Machado-Schiaffino, M. Olave, F. Henning, J. Torres-Dowdall, A. Härer, C. D. Hulsey, P. Franchini, Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**, 106–111 (2020).
33. C. F. Kratochwil, A. F. Kautt, A. Nater, A. Härer, Y. Liang, F. Henning, A. Meyer, An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes. *Nat. Commun.* **13**, 296 (2022).
34. R. L. Cosby, J. Judd, R. Zhang, A. Zhong, N. Garry, E. J. Pritham, C. Feschotte, Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* **371**, eabc6405 (2021).
35. J. Guo, G. H. Yue, J. Li, C. Wu, L. Wang, Y. Shen, Transposable element-mediated cis-regulation drives the evolution of *dmrt1* as a candidate master sex-determining gene in black carp. *Mol. Biol. Evol.* **42**, msaf239 (2025).
36. Y. H. Gray, It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**, 461–468 (2000).
37. N. Rafati, J. Chen, A. Herpin, M. E. Pettersson, F. Han, C. Feng, O. Wallerman, C.-J. Rubin, S. Péron, A. Cocco, Reconstruction of the birth of a male sex chromosome present in Atlantic herring. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24359–24368 (2020).
38. T. Sultana, A. Zamborlini, G. Cristofari, P. Lesage, Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* **18**, 292–308 (2017).
39. L.-F. Liang, S. M. Soyal, J. Dean, FIGα, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes. *Development* **124**, 4939–4947 (1997).
40. R. A. Bayne, S. J. Martins da Silva, R. A. Anderson, Increased expression of the FIGLA transcription factor is associated with primordial follicle formation in the human fetal ovary. *Mol. Hum. Reprod.* **10**, 373–381 (2004).
41. K. Wu, Y. Zhai, M. Qin, C. Zhao, N. Ai, J. He, W. Ge, Genetic evidence for differential functions of figla and nobox in zebrafish ovarian differentiation and folliculogenesis. *Commun. Biol.* **6**, 1185 (2023).
42. S. Yoshimoto, E. Okada, H. Umemoto, K. Tamura, Y. Uno, C. Nishida-Umehara, Y. Matsuda, N. Takamatsu, T. Shiba, M. Ito, A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2469–2474 (2008).
43. J. Zakany, D. Duboule, The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**, 359–366 (2007).
44. Z. Huang, T. Shi, B. Zheng, R. E. Yumul, X. Liu, C. You, Z. Gao, L. Xiao, X. Chen, APETALA 2 antagonizes the transcriptional activity of AGAMOUS in regulating floral stem cells in *Arabidopsis thaliana*. *New Phytol.* **215**, 1197–1209 (2017).
45. D. A. Gook, D. Edgar, J. Borg, M. Martic, Detection of zona pellucida proteins during human folliculogenesis. *Hum. Reprod.* **23**, 394–402 (2008).
46. D. Onichtchouk, K. Aduroja, H. G. Belting, L. Gnügge, W. Driever, Transgene driving GFP expression from the promoter of the zona pellucida gene zpc is expressed in oocytes and provides an early marker for gonad differentiation in zebrafish. *Dev. Dyn.* **228**, 393–404 (2003).
47. M. Matschiner, A. Böhne, F. Ronco, W. Salzburger, The genomic timeline of cichlid fish diversification across continents. *Nat. Commun.* **11**, 5895 (2020).
48. I. Irisarri, P. Singh, S. Koblmüller, J. Torres-Dowdall, F. Henning, P. Franchini, C. Fischer, A. R. Lemmon, E. M. Lemmon, G. G. Thallinger, Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* **9**, 3159 (2018).
49. F. Ronco, M. Matschiner, A. Böhne, A. Boila, H. H. Büscher, A. El Taher, A. Indermaur, M. Malinsky, V. Ricci, A. Kahmen, Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature* **589**, 76–81 (2021).
50. W. P. Maddison, Gene trees in species trees. *Syst. Biol.* **46**, 523–536 (1997).
51. J. H. Degnan, N. A. Rosenberg, Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340 (2009).
52. J. C. D. Hotopp, Horizontal gene transfer between bacteria and animals. *Trends Genet.* **27**, 157–163 (2011).
53. A. M. Ivancevic, R. D. Kortschak, T. Bertozzi, D. L. Adelson, Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* **19**, 85 (2018).
54. A. El Taher, F. Ronco, M. Matschiner, W. Salzburger, A. Böhne, Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. *Sci. Adv.* **7**, eabe8215 (2021).
55. L. Blumer, V. Burskaia, I. Artiushin, J. Saha, J. C. Garcia, F. C. Jiménez, A. Hooft van Huysdynen, J. Elkin, B. Fischer, N. Van Houtte, Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid radiation. *Science* **388**, eadr9961 (2025).
56. J. A. M. Graves, Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901–914 (2006).
57. X. Ma, L. a. Yu, M. Fatima, W. H. Wadlington, A. M. Hulse-Kemp, X. Zhang, S. Zhang, X. Xu, J. Wang, H. Huang, The spinach YY genome reveals sex chromosome evolution, domestication, and introgression history of the species. *Genome Biol.* **23**, 75 (2022).
58. D. Chalopin, J.-N. Volff, D. Galiana, J. L. Anderson, M. Schartl, Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Res.* **23**, 545–560 (2015).
59. J. Wang, J.-K. Na, Q. Yu, A. R. Gschwend, J. Han, F. Zeng, R. Aryal, R. VanBuren, J. E. Murray, W. Zhang, Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13710–13715 (2012).
60. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
61. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
62. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
63. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
64. O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
65. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
66. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
67. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–14.10.14 (2009).
68. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
69. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
70. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
71. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
72. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
73. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
74. N. A. Baird, P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, E. A. Johnson, Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* **3**, e3376 (2008).
75. J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
76. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
77. P. Rastas, Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726–3732 (2017).
78. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

79. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
80. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
81. D. M. Simmons, J. L. Arriza, L. Swanson, A complete protocol for in situ hybridization of messenger RNAs in brain and other tissues with radio-labeled single-stranded RNA probes. *J. Histotechnol.* **12**, 169–181 (1989).
82. F. Heigwer, G. Kerr, M. Boutros, E-CRISP: Fast CRISPR target site identification. *Nat. Methods* **11**, 122–123 (2014).
83. L. Wang, F. Sun, Z. Y. Wan, B. Ye, Y. Wen, H. Liu, Z. Yang, H. Pang, Z. Meng, B. Fan, Genomic basis of striking fin shapes and colors in the fighting fish. *Mol. Biol. Evol.* **38**, 3383–3396 (2021).
84. S. Knudsen, Promoter2.0: For the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356–361 (1999).
85. D. Brawand, C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
86. S. Chang, G. Ngoh, L. Kueh, Q. Qin, C. Chen, T. Lam, Y. Sin, Development of a tropical marine fish cell line from Asian seabass (*Lates calcarifer*) for virus isolation. *Aquaculture* **192**, 133–145 (2001).
87. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
88. S. E. Millar, E. Lader, L.-F. Liang, J. Dean, Oocyte-specific factors bind a conserved upstream sequence required for mouse zona pellucida promoter activity. *Mol. Cell. Biol.* **11**, 6197–6204 (1991).
89. D. E. Mold, A. E. Dinitz, D. R. Sambandan, Regulation of zebrafish zona pellucida gene activity in developing oocytes. *Biol. Reprod.* **81**, 101–110 (2009).
90. L. Wang, S. Liu, Y. Yang, Z. Meng, Z. Zhuang, Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers. *Mol. Ecol.* **31**, 5729–5744 (2022).
91. L. Wang, B. Bai, P. Liu, S. Q. Huang, Z. Y. Wan, E. Chua, B. Ye, G. H. Yue, Construction of high-resolution recombination maps in Asian seabass. *BMC Genomics* **18**, 63 (2017).
92. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
93. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
94. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
95. S. Kalyaanamoorthy, B. Q. Minh, T. K. Wong, A. Von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
96. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77–80 (2010).
97. Z. Yang, R. Nielsen, Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
98. V. Ranwez, E. J. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).

Acknowledgments

Funding: This research was supported by internal funding from the Temasek Life Sciences Laboratory, Singapore. **Author contributions:** G.H.Y., L.W., M.S., F.P., and A.M. conceived the study. G.H.Y., F.P., M.S., and A.M. supervised the whole study. L.W., F.S., Z.Y., M.L., J.W., S.Y., and Y.W. performed experiments. L.W., F.S., and Z.Y. analyzed the data. L.W. wrote the paper with input from the other authors. L.W., G.H.Y., M.S., F.P., and A.M. interpreted the findings in biological context and commented on the manuscript. All authors discussed the results and approved the final version of the paper. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials and are publicly available. Raw sequencing reads for genome assembly, RNA sequencing, and RAD sequencing are archived in the National Genomics Data Center of China National Centre for Bioinformation (NGDC; https://ngdc.cncb.ac.cn/) with BioProject accession PRJCA036437. Genome sequences are archived in the Genome Warehouse of NGDC with accession GWHFPXM00000000.1 and GWHFPXL00000000.1. This study did not generate any new materials.

Submitted 13 November 2025
Accepted 24 April 2026
Published 3 June 2026
10.1126/sciadv.aed6242

ScienceAdvances

Transposons drove the evolution of a male sex-determining gene from a female gene with deep introgression across cichlids

Le Wang, Fei Sun, Zituo Yang, May Lee, Joey Wong, Shadame Yeo, Yanfei Wen, Axel Meyer, Francesc Piferrer, Manfred Schartl, and Gen Hua Yue

Sci. Adv. **12** (23), eaed6242. DOI: 10.1126/sciadv.aed6242

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.