SPECIAL ISSUE

MOLECULAR ECOLOGY WILEY

Molecular parallelism in the evolution of a master sex-determining role for the anti-Mullerian hormone receptor 2 gene (*amhr2*) in Midas cichlids

Camila L. Nacif[1] [iD] | Claudius F. Kratochwil[2] [iD] | Andreas F. Kautt[2] [iD] |
Alexander Nater[2] [iD] | Gonzalo Machado-Schiaffino[2] [iD] | Axel Meyer[2] [iD] |
Frederico Henning[1,2] [iD]

[1]Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Cidade Universitária, Rio de Janeiro, Brazil

[2]Department of Biology, University of Konstanz, Konstanz, Germany

Correspondence
Frederico Henning, Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Cidade Universitária, Rio de Janeiro, Brazil.
Email: fhenning@acd.ufrj.br

Axel Meyer, Department of Biology, University of Konstanz, Konstanz, Germany.
Email: axel.meyer@uni-konstanz.de

Present address
Claudius F. Kratochwil, Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki, Finland

Andreas F. Kautt, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

Gonzalo Machado-Schiaffino, Department of Functional Biology, Area of Genetics, University of Oviedo, Oviedo, Spain

Abstract

The evolution of sex chromosomes and their differentiation from autosomes is a major event during genome evolution that happened many times in several lineages. The repeated evolution and lability of sex-determination mechanisms in fishes makes this a well-suited system to test for general patterns in evolution. According to current theory, differentiation is triggered by the suppression of recombination following the evolution of a new master sex-determining gene. However, the molecular mechanisms that establish recombination suppression are known from few examples, owing to the intrinsic difficulties of assembling sex-determining regions (SDRs). The development of forward-genetics and long-read sequencing have generated a wealth of data questioning central aspects of the current theory. Here, we demonstrate that sex in Midas cichlids is determined by an XY system, and identify and assemble the SDR by combining forward-genetics, long-read sequencing and optical mapping. We show how long-reads aid in the detection of artefacts in genotype–phenotype mapping that arise from incomplete genome assemblies. The male-specific region is restricted to a 100-kb segment on chromosome 4 that harbours transposable elements and a Y-specific duplicate of the anti-Mullerian receptor 2 gene, which has evolved master

sex-determining functions repeatedly. Our data suggest that *amhr2Y* originated by an interchromosomal translocation from chromosome 20 to 4 pre-dating the split of Midas and Flier cichlids. In the latter, it is pseudogenized and translocated to another chromosome. Duplication of anti-Mullerian genes is a common route to establishing new sex determiners, highlighting the role of molecular parallelism in the evolution of sex determination.

KEYWORDS
genetic mapping, GWAS, parallel evolution, sex chromosome evolution, XY sex determination

1 | INTRODUCTION

Uncovering universal mechanisms that underlie genomic and phenotypic diversification is one of the main goals in contemporary evolutionary biology (Blount et al., 2018; Lassig et al., 2017). Evolutionary convergence is ubiquitous (Blount et al., 2018) and there have been numerous suggestions that molecular parallelism in shared developmental programmes play an important role in determining evolutionary trajectories. However, generalities remain hard to identify (Elmer & Meyer, 2011). Because it is a genetically tractable trait and evolved repeatedly in some groups of organisms (Bachtrog et al., 2014), sex determination is an excellent trait to study the predictability of evolutionary processes and ask whether there are common mechanisms and routes towards the evolution of similar phenotypes. While macroevolutionary patterns have typically been seen as contingent since Gould's influential book "*Wonderful Life*" (Gould, 1989), recent studies show that evolution reuses molecular routes and is predictable at some level (Kratochwil et al., 2019). This seems particularly so in the establishment of sex determination, where certain genes and molecular mechanisms are repeatedly involved (Feron et al., 2020; Graves & Peichel, 2010).

Although sex-determining genomic regions are typically inaccessible due to the association with complex repeats, recent technological advances now make it possible to efficiently identify and dissect the sex-determining regions (SDRs) of nonmodel organisms (e.g., Peichel et al., 2020). The evolution of sex determination is a major event in genome evolution, causing large-scale changes in chromosome morphology and gene content. The accumulation of repetitive elements makes the genomic assembly of SDRs extremely challenging (Miga et al., 2020; Peichel et al., 2020; Sedlazeck et al., 2018). However, the discovery and assembly of SDRs has been facilitated recently due to advances in long-read sequencing technologies (Palmer et al., 2019; Sember et al., 2021). Long-read methodologies are especially useful to assemble SDRs, as they generate scaffolds spanning complex genomic regions (Fraser et al., 2020; Nakamoto et al., 2021; Peichel et al., 2020; Qu et al., 2021; Tao, Xu, et al., 2021; Xue et al., 2021).

The classic theory of sex chromosome evolution posits that the expansion of the nonrecombining region, degeneration of the heterogametic chromosome and expansion of repetitive sequences result predictably from well-established genetic mechanisms such as Muller's ratchet and the Hill-Robertson effect (Charlesworth, 1991). The straightforward nature of the genetic mechanisms involved in the evolution of sex chromosomes led to the perception that heteromorphic chromosomes are an inevitable outcome following the origin of a new master sex-determining gene (Kratochvíl et al., 2021; Vicoso, 2019). The increase in genomic data for nonmodel species has challenged the perception that highly differentiated sex chromosomes are inevitable (Chalopin et al., 2015; Furman et al., 2020). For example, sexual systems and mechanisms have evolved repeatedly in fish, amphibians and reptiles (Bachtrog et al., 2014; Devlin & Nagahama, 2002), where they undergo frequent turnovers and rarely achieve high levels of differentiation (Palmer et al., 2019; Vicoso, 2019).

Teleosts represent more than half of the extant species of vertebrates and exhibit great diversity, and fast turnover in sex determination and differentiation mechanisms (Ashman et al., 2014; Bachtrog et al., 2014), including environmental as well as genetic mechanisms of various types (XY, ZW and multiple SD) (Kuwamura et al., 2020; Pennell et al., 2018). In teleosts, sex chromosomes are often homomorphic; that is, the region in which gene content and repetitive elements differ is restricted to a small section that is undetected using most cytogenetic techniques (Bachtrog et al., 2014; Schartl et al., 2016). The wide spectrum of sex-determining mechanisms found among closely related species (Nagahama et al., 2021) makes fishes well-suited models for studying the initial steps during sex chromosome evolution and also to test for molecular parallelism in the origin of master sex-determining (MSD) genes (Böhne et al., 2019; Godwin & Roberts, 2018). Several MSD genes have already been identified in teleosts, including *dmrt1Y* (*Oryzias latipes*), *sox3* (*Oryzias dancena*), *gsdfy* (*Oryzias luzonensis*), *amhy* (*Oreochromis niloticus*) and *amhr2Y* (*Takifugu rubripes*, *Perca flavecens*, *Percoglossus altivelis*, *Phyllopteryx taeniolatus* and *Syngnathoides biaculeatus*) (Duan et al., 2021; Feron et al., 2020; Gao et al., 2020; Ieda et al., 2018; Kamiya et al., 2012; Li et al., 2015; Myosho et al., 2012; Nakamoto et al., 2021; Nanda et al., 2002; Qu et al., 2021; Takehana et al., 2014). Although the variety of MSD genes is daunting, the repeated involvement of many of these suggests a large degree of molecular parallelism and that the evolution of novel sex-determiners is predictable.

Cichlids are prime examples of explosive adaptive radiations, in which flocks of recently diverged and phenotypically diverse species are suitable for the genetic dissection of numerous ecologically

relevant traits (Henning & Meyer, 2014). An array of sexual chromosomes and sex determination mechanisms have been described recently in cichlids (Böhne et al., 2019; Gammerdinger & Kocher, 2018; Godwin & Roberts, 2018), but identification of sex-determining genes has proven elusive. More than 15 different genomic regions have been associated with sex determination, with numerous instances of sex chromosome turnover described in this lineage (El Taher et al., 2020; Gammerdinger & Kocher, 2018; Lichilín et al., 2021; Tao, Xu, et al., 2021; Vicoso, 2019). Nicaraguan Midas cichlids comprise an adaptive radiation of over 10 species belonging to the Central American genus *Amphilophus*, including some of the best-known examples of ecological sympatric divergence (Barluenga et al., 2006). More than 10 species of Midas cichlids have already been described, many of which are endemic to crater lakes (see figure 1 in Kautt et al., 2021 for a graphical summary and Table S1 in the present study). Crater lake species originated from source populations of *Amphilophus citrinellus* and *A. labiatus* inhabiting the Great Lakes Nicaragua and Masaya. Between-species genetic divergence is low in this system, and extensive sharing of genetic variation for several traits of ecological relevance across the different crater lake assemblages has been documented (Kautt et al., 2020). Earlier behavioural work proposed that sex in the Midas cichlids is socially determined by agonistic social interactions based on body size (Francis & Barlow, 1993). However, this was later disputed and it was shown that the larger size in males was the result rather than the cause of sex determination (Oldfield, 2009).

Here we combine forward genetics (linkage and genome-wide association [GWA] mapping), long-read sequencing and optical mapping to dissect and characterize the Y-specific region of Midas cichlids. We show that sex is controlled by an XY system, probably due to a male-specific duplicate of the *amhr2* gene on chromosome 4 that evolved via interchromosomal translocation from chromosome 20. The data suggest that the Midas cichlid sex chromosome is at least 20-30 million years old and has undergone little differentiation in this period.

2 | MATERIALS AND METHODS

2.1 | Linkage mapping

To locate the SDR using cross-based linkage-mapping, a panel of 208 F_2s was established by crossing a male *Amphilophus labiatus* and a female *A. astorqui*, both wild-caught from Crater Lakes Masaya and Xiloá, respectively. From this cross, we generated an F_1 population, from which a male and female were selected and intercrossed to generate the F_2 generation. Due to the low level of reproductive isolation, interspecies crosses in Midas cichlids are readily obtained, providing an optimal segregation panel for genetic mapping that maximizes divergence in both genetic markers and phenotypic traits.

All fish were raised to maturity controlling for the social environment (i.e., maintaining equal tank densities and male–female ratios). F_2s were sexed by the direct observation of pair formation and

mating, resulting in phenotypic data for 207 F_2s, of which 111 were females and 96 were male. double digest restriction-site associated DNA (ddRAD) libraries were prepared as described by Recknagel et al. (2013). Data was processed in STACKS using the publicly available female Midas cichlid genome (NCBI accession ASM1343575v1) as a reference. The RAD catalogue was coded manually, in order to use all possible marker segregation types to build both uni- and biparental maps using JOINMAP version 5. These include markers that are homozygous for different alleles in both parentals (AAxBB) and also other segregation types that can be used depending on the allele that is transmitted by the F_1s. The parental origin of the "A" allele can be ambiguous when parentals share an allele (AAxAB or, conversely, ABXBB), but these markers can be used provided that both F_1s inherit the A allele from the heterozygous parental.

A genome-wide linkage map was constructed using the parameters and filters that were previously described in detail elsewhere (Henning et al., 2017). Trait mapping was performed using the QTL package in R 4.1.1 (R Core Team, 2020). To first identify linkage to sex on chromosome 4, we built a genome-wide F_2 biparental map. We used F_1 uniparental maps of chromosome 4 to distinguish between female and male heterogamety (XY or ZW) and to identify homogametic and heterogametic F_2s for recombinant breakpoint mapping. By selecting markers that were homozygous in one F_1, but heterozygous in the other (e.g., AAxAB or ABXBB markers), we tracked each copy of chromosome 4 that was transmitted to the F_2s (see File S1). This is not possible using F_2 maps using fixed differences between the parentals (i.e., AAxBB markers) because the expected proportion of heterozygotes is the same regardless of the system being ZW or XY: under full association, the expected ratio is 1:1:0 and 0:1:1 for AA:AB:BB genotypes in males and females, respectively.

2.2 | Long-read sequencing and optical mapping

Following the discovery of an XY system on chromosome 4, that the causal locus was shared among different Midas populations and that the female reference genome was probably lacking a Y-specific interval of unknown size, we generated a new reference Y chromosome by combining PacBio HiFi and Bionano optical maps. Whole blood was drawn using blood collection tubes coated with EDTA (Sarstedt) from the caudal vein of an F_4 individual obtained from a cross between a golden *A. citrinellus* (Crater Lake Masaya) and a dark *A. citrinellus* (Great Lake Nicaragua). High-molecular-weight DNA was isolated using the SP Blood & Cell DNA Isolation kit (Bionano). The sample was labelled using a DLS Labeling kit (Bionano) and was run on two flow cells generating a total of 4339 Gb of cmap data. This individual was chosen for sequencing to maximize causal variant discovery and simultaneously led to the identification of the casual variant underlying the colour polymorphism after which these fish are named (Kratochwil et al., 2022).

Accurate long-read sequencing data (PacBio HiFi or CCS) were obtained from the same blood sample and were concomitantly used in Kratochwil et al. (2022). Briefly, DNA was sheared to 15–23 kb

and sequencing libraries were prepared using the SMRTbell Express Template Prep Kit 2.0, followed by size selection on the SageELF (SAGE Science). Sequence data were collected using a single SMRT cell on the PacBio Sequel II system using HiFi sequencing protocols and Sequencing kit V2 over 30 h. Raw PacBio sequences were processed in the CCS software to generate ~20 Gb of HiFi reads. Bionano and PacBio data were collected at Praxis Genomics LLC.

2.3 | Assembly of the Midas male-specific region

HiFi reads were de novo assembled using the recently developed haplotype-aware assembler HIFIASM (Cheng et al., 2021). All optical mapping analyses was performed in BIONANO ACCESS (version 1.6.1) using the BIONANO SOLVE software (version 3.6) through "compute on demand," apart from the isolation of Y and X contigs, which was done by manually editing the cmap files. Optical maps were assembled de novo using both the haplotypic and nonhaplotypic pipelines and using the primary HiFi assembly as reference. The de novo assembly pipeline implemented in BIONANO SOLVE is not reference-guided, but including it allows for the straightforward identification through a table of correspondence between optical and sequence data supplied by the software. We identified the de novo assembled Bionano cmaps that corresponded to the male-specific region by supplying a reference that contained the primary assembly supplemented with a 300-kb contig that included the y-specific interval. Therefore, we could readily obtain the cmaps that corresponded to both the X and Y scaffolds. We then manually retrieved the corresponding cmaps and performed hybrid-scaffolding separately for the two sets corresponding to the Y and X haplotypes.

To generate a chromosome 4 assembly that included the Y-specific region (YSR), hybrid-scaffolding was performed on a subset of data containing only those optical maps and contigs that corresponded to chromosome 4 and the YSR. As haplotypic hybrid-scaffolding is currently unsupported in BIONANO ACCESS, we screened the primary contigs for sequences originating from chromosome 4, as well as the alternative contigs (i.e., haplotigs) and sequences originating from the X and Y chromosomes. To generate the remaining 23 chromosomes of the reference male genome used for association mapping and variant screening, we used the primary contigs generated by HIFIASM, which were then scaffolded with a haplotype-unaware de novo assembly of optical maps using the assembly and HS pipelines in BIONANO ACCESS (allowing us to cut both contig sequences and optical maps). The chromosome 4 assembly that resulted from this whole genome approach corresponds to the X chromosome, which was replaced with the Y chromosome for GWA and copy number variation (CNV) detection. Super-scaffolds were filtered by size based on the N99 value. We used RAGTAG (Alonge et al., 2019) without query editing to place and orientate the super-scaffolds on chromosomes based on the *A. citrinellus* reference genome (Kautt et al., 2020). We evaluated the quality of the genome assembly using the GENOMEQC online server (Manchanda et al., 2020). Finally, we named the male haplotig assembly chromosomes based

on the results of whole-genome alignment to the *A. citrinellus* reference assembly (GenBank accession no.: ASM1343575v1) using MINIMAP2 (Li, 2018).

2.4 | GWA and CNV detection

To fine-map and screen for causal variants, we used the recently published short read data covering all Midas cichlid species and ecomorphs (Kautt et al., 2020) mapped against our newly assembled male genome. Roughly 200 of the samples from Kautt et al. (2020) were phenotyped for external sexual morphology during sampling and could be used to test for genotype–phenotype associations ($n = 197$, 113 males and 87 females). Variants and genotype calling was performed with FREEBAYES. Mapping coverage was estimated by SAMTOOLS DEPTH (Li et al., 2009) and the mapping result visually inspected with INTEGRATIVE GENOMICS VIEWER version 2.3.68 (IGV) (Robinson et al., 2011) for regions of interest. Read mapping, variant and genotyping calling were conducted using the scripts made publicly available by Kautt et al. (2020) (https://github.com/alexnater/midas-genomics). Phenotype–genotype associations were tested using the dominant generalized linear model (GLM) implemented in PLINK2 (Chang et al., 2015), using marker data that had a minimum genotyping frequency of 80% and minor allele frequencies above 5%. We filtered the association result for the 1000 most-significant p-values for each chromosome and plotted using the *manhattan* function of the QQMAN R package (Turner, 2018). These analyses were also performed with the same pipeline against the female reference genome.

Regions with significant peaks of association, as defined by inspecting the distribution of p-values with a quantile–quantile plot (R/QQMAN), were subject to further analysis. We screened for differences in depth of coverage between males and females in these regions using *generate_regions.py* in FREEBAYES (Garrison & Marth, 2012) and SAMTOOLS DEPTH (Li et al., 2009), extracting the depth of mapping coverage every 1 kb, with the minimum quality (MAPQ) for mapped base and read of 20 and 30, respectively. Mapping results were visually examined using IGV (Robinson et al., 2011).

2.5 | Annotation of the YSR and candidate gene identification

Annotation of the Midas cichlid YSR was performed using BLASTX and BLAT in GenBank and ENSEMBL, respectively. Transposable elements (TEs) were annotated using the tools available in FishDB (Yang et al., 2020). Following the haplotypic assembly of the Midas cichlid Y-chromosome, we performed BLASTX searches across the 300-kb section containing the YSR and neighbouring highly divergent sequences. Maximum likelihood gene-trees were used to identify the causal gene and conserved synteny was used to annotate the surrounding region. Because initial BLASTX searches matched the Bmpr2-like protein, we conducted a broad analysis including the

closest members of this gene family (*amhr2* and *bmpr2*) according to the literature (Adolfi et al., 2019; Massagué et al., 1994).

Following the identification of our candidate as an anti-Mullerian receptor 2 gene (*amhr2*), we retrieved further sequences in order to phylogenetically place the duplication events, which included cases in which *amhr2* duplications are known, as well as the closest related available complete genome, the Flier cichlid (*Archocentrus centrarchus*) (Rhie et al., 2021; Xiong et al., 2021). Sequences from representative fish and vertebrate groups were retrieved from EMSEMBL version 104 and GenBank. Besides the protein sequences from Midas and Flier cichlids, we also added sequences of the MSD gene from *T. rubripes* and *P. flavescens* (Table S1). Initial searches on annotated sequences failed to locate an *amhr2Y* homologue in the *Ar. centrarchus* genome, which prompted us to perform a local BLASTN search using the entire Midas cichlid coding sequence (CDS) as query.

Protein sequences were aligned using MAFFT (Katoh & Standley, 2013) with the *mafft-linsi* setting and were trimmed using TRIMAL (Capella-Gutiérrez et al., 2009) using -gt 0.6 -st 0.005. JMODELTEST2 (Darriba et al., 2012) was used to select the substitution model (JTT+I+G4). Phylogenetic inference with 10,000 bootstrap replicates was performed in IQTREE2 (Minh et al., 2020).

2.6 | Comparative genomics

Large genomic sections and entire assemblies were aligned using MINIMAP2, and visualized through dotplots using the D-genies web interface (Cabanettes & Klopp, 2018). To determine the extent of the region that originally translocated from chromosome 20 to 4, we extracted 2 Mb up- and downstream of *amhr2* and *amhr2Y* from the Midas cichlid and other teleost fish species: *Ar. centrarchus*, Nile tilapia (*Oreochromis niloticus*), southern platyfish (*Xiphophorus maculatus*), medaka (*Oryzias latipes*) and fugu (*Takifugu rubripes*). Synteny of the genomic region harbouring *amhr2* genes was analysed using sequences from the same taxa listed above. Genes present in each extracted region were manually inspected using BLASTN and BLASTX searches. Orthologous genes were grouped by best reciprocal hits using predicted protein sequences. The accession numbers of all the sequences are given in Table S1.

To test whether the fragment of *amhr2Y* from *Ar. centrarchus* plays a sex-determining role, we screened for coverage variation using whole-genome sequencing (WGS) Illumina short-read data from 20 unsexed, wild-caught individuals available from Xiong et al. (2020).

3 | RESULTS

3.1 | Linkage mapping

Linkage mapping identified a single interval on chromosome 4 that segregates with the sex phenotype in the F_2 panel (Figure 1a). Due to complete linkage of marker and phenotype data, we could narrow

the region to 700 kb by recombinant breakpoint analysis (Figure 1b). Uniparental maps showed no differences in recombination frequency between the male and female F_1s at the SDR (Figure S1). Phasing of parental X and Y chromosomes based on F_1 segregation showed that the males are the heterogametic sex (File S1).

3.2 | Y chromosome assembly

The HIFIASM primary assembly contained a large segment of chromosome 4 representing the X chromosome (6 Mb) and a smaller contig containing the YSR (306 kb). Screening the haplotigs led to the identification of a 1.5-Mb contig containing the YSR and flanking sequence (Figure 2a). Due to the presence of large repetitive regions at the ends of this haplotig, it could not confidently be assembled to the remaining chromosome 4 contigs. Integration with optical maps made it possible to experimentally join this 1.5 Mb to the remaining chromosome 4 contigs and validate the repeat structure found in the haplotigs (Figure 2b).

The de novo genome based on HiFi data alone and assembled with HIFIASM yielded a primary assembly of 950 Mb containing 1028 contigs and a contig N50 of 2.83 Mb. Hybrid-scaffolding with Bionano optical maps reduced the total size to 907 Mb, comprising 131 super-scaffolds (Figure S2). After this step, small super-scaffolds were ~170 kb and were filtered out. After filtering by N99, 102 super-scaffolds were retained and after reference-based scaffolding, the 24 chromosomes were assembled, and the genome was reduced to 894 Mb (Table 1; Figure S3). This assembly is consistent with the expected genome size of the *Amphilophus citrinellus* (ASM1343575v1) and *Ar. centrarchus* (GCA_007364275.2:933 Mb) assemblies available in GenBank.

3.3 | GWA mapping and causal variant detection

Genome-wide short-read (WGS) data from 112 males and 87 females, from the Great Lake Nicaragua and Managua, and Crater Lakes As. León, As. Managua Apoyeque, Apoyo, Masaya, Tiscapa and Xiloá, were mapped to our male genome assembly. A total of 5,816,655 single nucleotide polymorphisms (SNPs) were detected in the variant calling step, and after normalization and filtering 2,600,544 high-quality SNPs were retained. A single peak of phenotype–genotype association was detected on chromosome 4 (Figure 1c). Screening for causal variants led to the identification of a region present only in males which we refer to subsequently as the Midas cichlid YSR (Figure 1d).

3.4 | Annotation of the Midas cichlid YSR

The Midas cichlid YSR and spans ~100 kb (30.67–30.78 Mb) on chromosome 4 (Figure S1) and has a total length of 36 Mb. The YSR contains a single complete coding gene (*amhr2Y*, see Section 3.5), a few partial
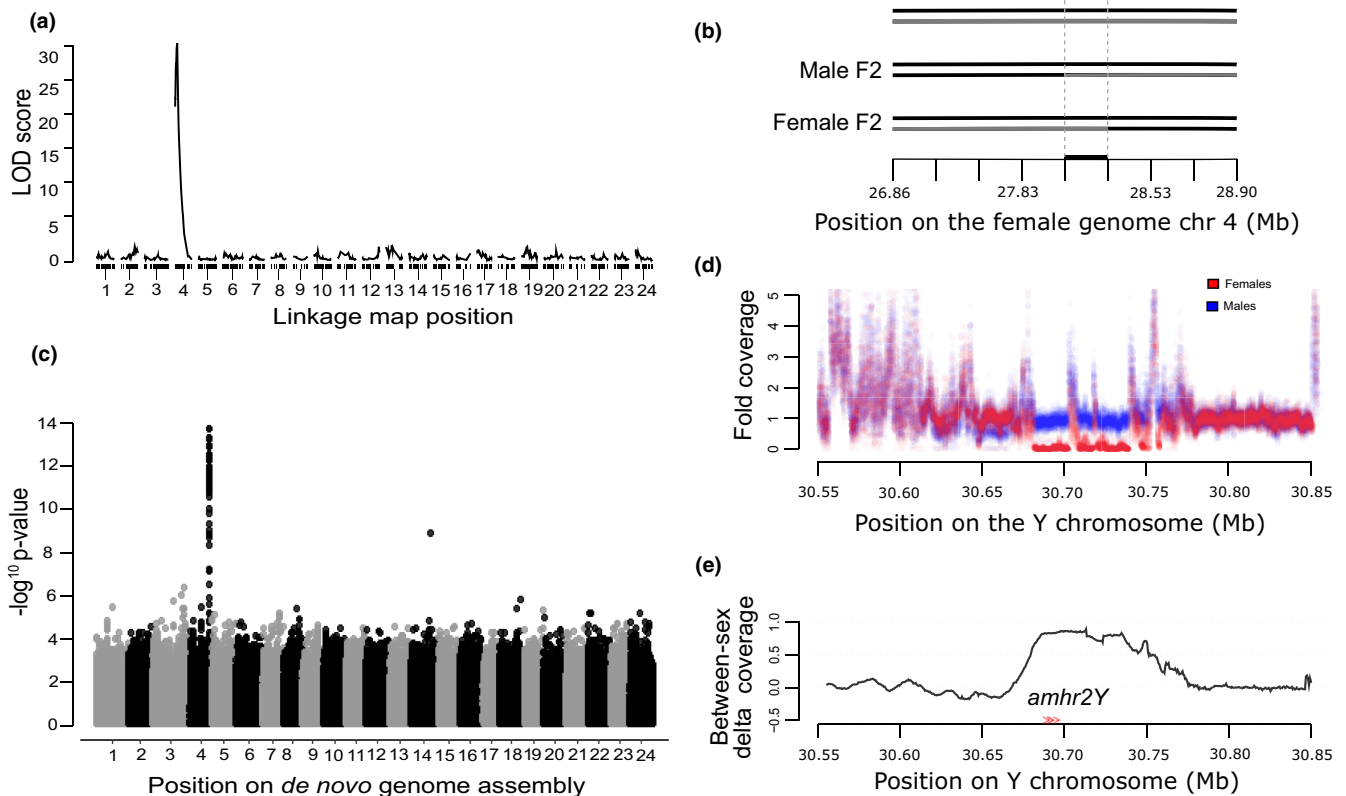
FIGURE 1 Forward-genetic mapping of sex determination in Midas cichlids. (a) Linkage mapping results in a single peak of association between sex and genotype on chromosome 4. (b) Recombination breakpoint analysis narrows the candidate region to a 700 kb interval on chromosome 4 of the Midas cichlid reference genome available from NCBI. Further analysis indicated male heterogamety (XY). (c) Marker–phenotype association in natural populations using the newly assembled male genome confirms and fine-maps the previous interval. (d) Read mapping across 300 kb centred at the male-specific region. A region of ~100 kb is present only in males and absent in females, forming blue and red lines at 1× and 0× coverage region indicated. (e) The same region as in (d), but depicted as the difference between the mean fold coverage of males and females to remove noise from neighbouring repetitive regions

genes (*usp43*, *znf72l* and *klhl10*), and many TEs (Figure 2). The YSR is not located within an inversion and does not seem to be significantly expanded. We detected a 500-kb section upstream of the YSR that is composed of large tandem repeats (repeat unit size = 40 kb). Within each repeat unit, BLASTX identified a copy of the *sco1* gene flanked by two copies of an unannotated TE in direct orientation. While we initially suspected that this was a chimeric assembly artefact, it was validated using optical maps (Figure 2b). Mapping coverage using WGS data revealed considerable CNV from 0 to 6 copies in population samples which varied regardless of sex (Figure S4). An accumulation of LTR- and LINE-type transposons is located upstream of the YSR (positions 30.36–30.67 Mb).

3.5 | Candidate-gene identification and comparative genomics

Annotation of the candidate genes using BLASTX resulted in top hits with members of related gene-families (e.g., *bmpr2*-like), indicating that the publicly available automatic annotations did not reflect orthologous relations and had to be manually curated. Phylogenetic trees including peptide sequences of Amhr2 and Bmpr2 resulted in

the clear nesting of our candidate gene within the *amhr2* gene tree (Figure S5). The duplication node clearly pre-dates the *Amphilophus–Archocentrus* split (Figure 3a). The *amhr2* gene is present on Midas cichlid chromosome 20, while *amhr2Y* is present on chromosome 4. Alignments of the autosomal (*amhr2*) and the sex-related copy (*amhr2Y*) revealed that the gene structure (11 exons) and the open-reading frame is conserved, but that *amhr2Y* has accumulated a large number of amino acid substitutions (Figure 3c).

A sequence orthologous to the Midas *amhr2Y* was found on chromosome 6 of *Ar. centrarchus* which contains only the last two exons (exons 10 and 11) and, also, a premature stop codon in exon 11. Conserved synteny with neighbouring genes further confirmed the orthologous relationship of the Midas and Flier cichlid (*Ar. centrarchus*) *amhr2* with other teleosts. It also indicated an absence of conserved synteny between the Midas and Flier cichlid *amhr2Y* orthologues. These data suggest that this copy is pseudogenized and suffered a secondary translocation. To test for the presence of *amhr2Y* and to confirm that this copy is nonfunctional, we mapped short reads from 20 unsexed individuals (Xiong et al., 2021) against the region harbouring the *amhr2Y* locus in the Midas cichlid male genome. We found that exons 10 and 11 were present in all the 20 individuals with the same

copy number. If these exons had sex-determination functions, there would be at least some individuals lacking this copy unless our sample consisted exclusively of males, which is highly unlikely (Figure S6).
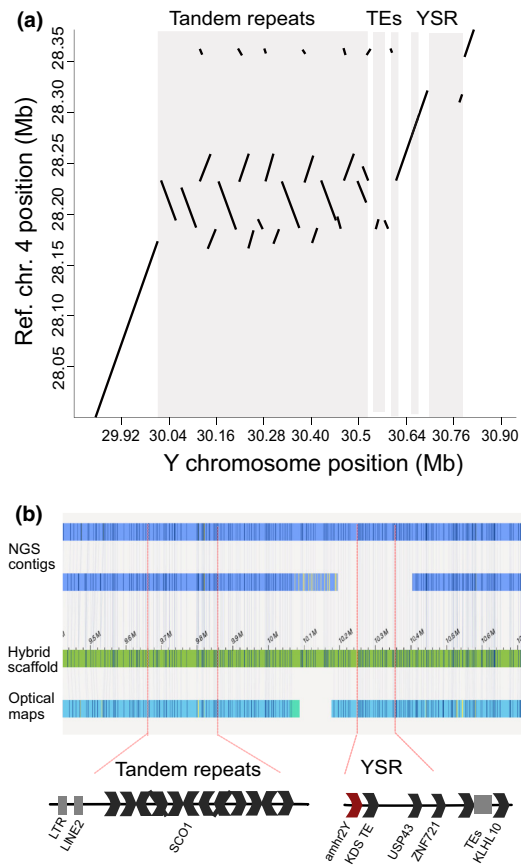
FIGURE 2 Genomic structure of the Y-specific region and neighbouring repeats on the Midas cichlid chromosome 4. (a) Dotplot of the alignment of the reference genome chromosome 4 (GenBank accession CM024195.1) and the Y chromosome assembled in the present study. (b) Hybrid-scaffolding of Bionano optical maps and HiFi contigs. The Y-haplotig is displayed on top. Primary assembly contigs from chromosome 4 are shown in the second row from the top
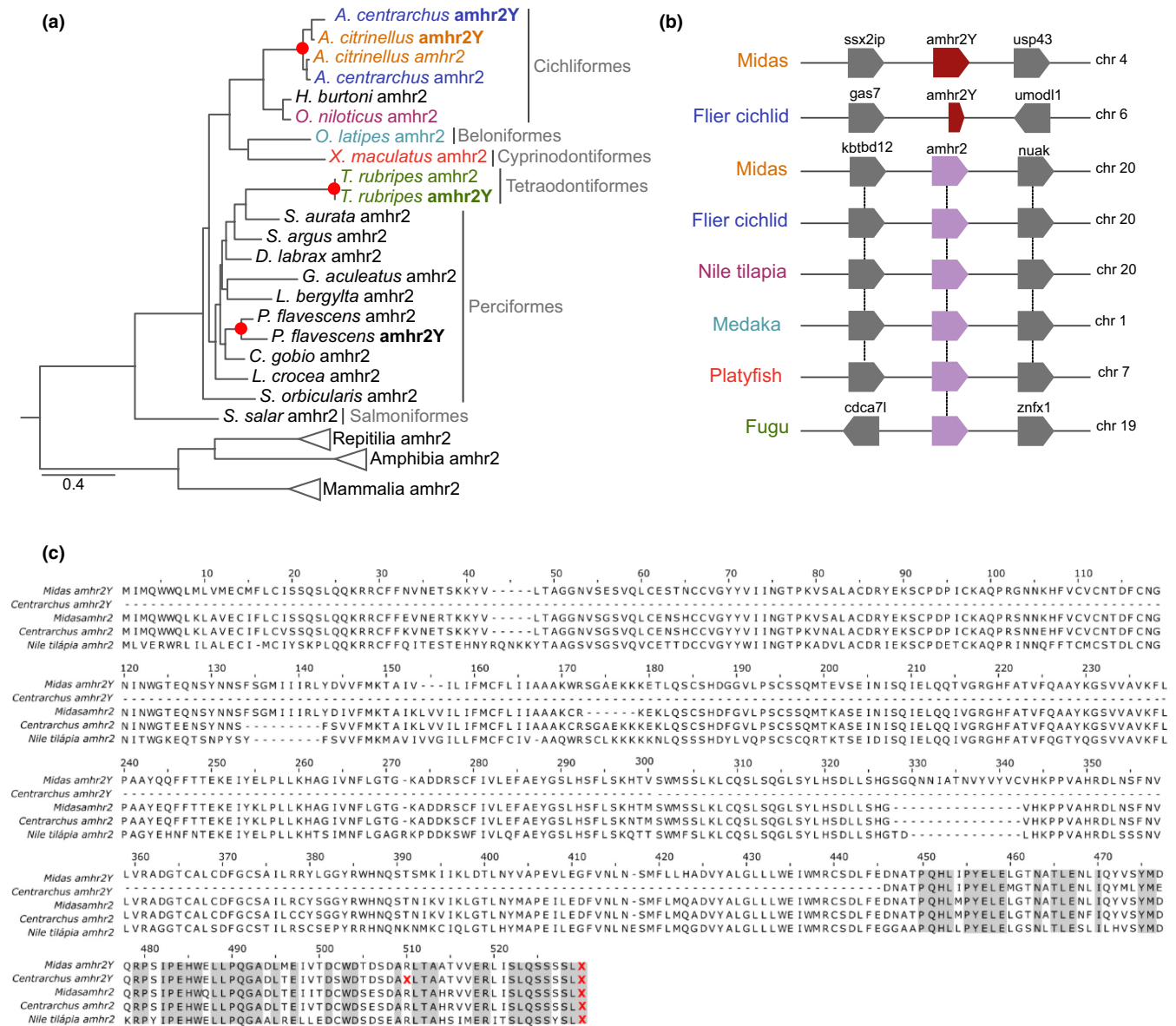
4 | DISCUSSION

Here we show that sex in the Midas cichlids is determined genetically based on a XY system with the probable molecular cause being a Y-specific duplicate of the anti-Mullerian receptor-2 gene (*amhr2Y*), a preferred route for the establishment of MSD genes in fishes. This gene functions as a receptor to the anti-Mullerian hormone, which induces Mullerian duct regression in males. Although teleosts lack Mullerian ducts, the effects are similar in that expression of *amh/amhr2* increases during puberty in males. The function of the *amh/amhr2* path in teleosts is in the maintenance and proliferation of germ cells (Adolfi et al., 2019).

The duplication and origin of *amhr2Y* pre-dates the split between *Amphilophus* and *Archocentrus*, where it has undergone a turnover and secondary rearrangement. Y-specific copies of *amhr2* have emerged at least five times independently as the MSD gene in teleosts (Kamiya et al., 2012; Nakamoto et al., 2021; Qu et al., 2021; Feron et al., 2020; see Table S2). We successfully generated a high-contiguity haplotypic assembly of a male and assembled the complete YSR using a combination of PacBio HiFi and Bionano optical maps. The YSR has many hallmarks of SDRs, including the accumulation of repetitive elements, but notably lacks inversions and has persisted for a long period of time without significant expansion of the nonrecombining region. Finally, our findings highlight the power of long-read sequencing to uncover complex causal variants and also points to issues that arise with genetic mapping based on short reads.

4.1 | Molecular parallelism in the repeated evolution of MSD genes

In birds and mammals, after the Y/W chromosomes undergo extensive degeneration and reach gene dosage compensation, turnover events become rare, which makes highly differentiated chromosomes a barrier to turnover (Bachtrog et al., 2014; Vicoso, 2019). Fish sex chromosomes, on the other hand, are often surprisingly

TABLE 1 Comparison of assembly statistics between the female reference genome available in GenBank (ASM1343575v1) and the male genome generated here; values are given in base pairs, unless noted otherwise

Assembly metrics	Male assembly (XY)	Reference genome (XX)
Number of scaffolds	24	24
Total size of scaffolds	915,194,499	887,751,838
Useful amount of scaffold sequences (≥25k nt)	915,194,499	887,751,838
Longest scaffold	61,990,358	62,521,812
Shortest scaffold	27,990,456	25,990,339
L50	11	12
%A	29.25	29.42
%C	20.36	20.38
%G	20.25	20.37
%T	29.31	29.44
Total number of Ns	7,617,186	3,516,008
%N	0.83	0.40

FIGURE 3 Identification of the putative sex-determining gene and timing of the Y-duplication. (a) Gene-tree of *amhr2* conducted using maximum likelihood. Circles denote duplication nodes. (b) Comparison of gene-synteny in the vicinity of the *amhr2* and *amhr2Y*. Vertical lines connect orthologues. (c) Alignment of Amhr2Y and Amhr2 orthologous amino acid sequences of Midas cichlid (*Amphilophus citrinellus*), Flier cichlid (*Archocentrus centrarchus*) and Nile tilapia (*Oreochromis niloticus*)

short-lived and have a high frequency of turnovers (El Taher et al., 2020; Ieda et al., 2018; Kitano & Peichel, 2012), which switches the master-determination role among genes in the above-mentioned pathways and restarts the sex chromosome differentiation cascade (Sember et al., 2021). The causes of sexual turnover are still debated and can occur through several mechanisms (Bachtrog et al., 2014; Blaser et al., 2013; El Taher et al., 2020; Kitano & Peichel, 2012; Palmer et al., 2019; Tao et al., 2021; Vicoso, 2019). Turnovers are frequent in cold-blooded vertebrate lineages that have homomorphic sex chromosomes (El Taher et al., 2020; Palmer et al., 2019; Pennell et al., 2018; Tao, Xu, et al., 2021; Vicoso, 2019).

Despite the exceptional variety of sex-determining genes and mechanisms in fishes, there seem to be preferential routes both

at the level of molecular mechanisms and also of the genes involved. The overwhelming majority of masculinizing genes in teleosts belong to the transforming growth factor beta (*TGF-β*) pathway (Table S2). This includes (i) *amhy*, the ligand of *amhr2* that determines sex in Nile tilapia (*O. niloticus*), pejerrey (*Odontesthes bonariensis*), northern pike (*Esox lucius*) and three-spined stickleback (*Gasterosteus aculeatus*); (ii) *gdf6* in turquoise killifish (*Nothobranchius furzeri*); (iii) *gsdf* in sablefish (*Anoplopoma fimbria*); (iv) *bmpr11b* in Atlantic herring (*Clupea harengus*); and (v) *amhr2y*, which has evolved the MSD function in yellow perch (*Perca flavescens*), common sea dragon (*Phyllopteryx taeniolatus*), alligator pipefish (*Syngnathoides biaculeatus*), ayu (*Plecoglossus altivelis*), and takifugu (*Takifugu rubripes*; *Takifugu pardalis*; *Takifugu poecilonotus*;

Takifugu obscurus) (Duan et al., 2021; Feron et al., 2020; Gao et al., 2020; Hattori et al., 2012; Herpin et al., 2021; Kamiya et al., 2012; Nakamoto et al., 2021; Pan et al., 2019; Peichel et al., 2020; Qu et al., 2021; Rafit et al., 2020; Reichwald et al., 2015).

The phylogenetic placement of duplication events (Figure 3a) clearly shows that *amhr2* evolved MSD function through gene duplication several times independently. Gene duplication appears to be the most frequent mechanism leading to turnover of MSD genes (Pan et al., 2019; Qu et al., 2021). The preponderance of gene duplicates in comparison to allelic variants in new sex-determining genes might be explained by the immediate establishment of a nonrecombining region when the duplicate is moved to another chromosome (Pan et al., 2019).

The possible molecular mechanisms are either two-step (duplication followed by translocation) or one-step (transposition or interchromosomal insertional translocation—IIT). Two-step mechanisms are reported in the literature (e.g., Furman et al., 2020; Pan et al., 2019). However, IITs are also compatible with our data and could explain the origin of the Midas cichlid *amhr2Y* more parsimoniously. Reports of spontaneous IITs are frequent in the human medical genetics literature (e.g., Nowakowska et al., 2012). While in the two-step mechanism two mutations are required, the one-step mechanism could automatically lead to a new MSD gene since an insertional translocation itself generates a duplication in F_1, depending on meiotic segregation.

As with chromosomal rearrangements in general, IITs stem from repeat-mediated nonhomologous exchange. An IIT event generates two types of balanced and two types of unbalanced gametes in equal proportions. Those are gametes (i) bearing a single copy on the ancestral chromosome (i.e., "wild type"); (ii) gametes bearing a single copy on the rearranged chromosome; (iii) gametes lacking the gene altogether; and (iv) those that possess two copies, one on each chromosome. In cases in which an additional copy of the translocated sex-related gene determines sex through dosage effects (e.g. Hattori et al., 2012), the fourth gamete type would be expected to develop as male and transmit both X and Y chromosomes in a single mutational step. The preponderance of the two-step mechanism would predict that chromosomes harbouring potential MSD-genes (e.g., *TGF-β* genes) are predisposed to becoming sex chromosomes, which has received limited empirical support (Kratochvil et al., 2021). The mixed support suggests a role for one-step mechanisms, but could also reflect the sheer number of potential MSD genes. A third, and less-favoured, mechanism is the direct transposition through the association to particular types of TEs. IITs seem consistent with the Midas cichlid case, but further work is necessary to distinguish between these three mechanisms as well as investigate the cause of MSD gene turnover in *Ar. centrarchus*.

4.2 | The Midas cichlid YSR is small and suppression of recombination is localized

Our data are consistent with a scenario in which a translocation occurred between chromosomes 4 and 20 before the split of *Amphilophus* and *Archocentrus*. Because the current estimates of divergence time are disparate, we used 30 million years, based on the mean of several studies available in www.timetree.org (Kumar et al., 2017). The Midas YSR is rather small (roughly 100 kb) and underwent limited differentiation given its age (20-30 million years). The absence of a wide region accumulating differentiation is consistent with our uniparental linkage data, as well as the mapping coverage analyses. This seems at odds with theoretical expectations and empirical examples such as stickleback, where the effects of recombination suppression led to the formation of various differentiation strata across a 17.5-Mb nonrecombining region (Peichel et al., 2020). Theory predicts the expansion of the nonrecombining region over time; however, counter examples such as the present one in Midas cichlids have accumulated recently, supporting the perception that current theory does not fully explain sex chromosome evolution in most organisms (Vicoso, 2019). For instance, in northern pikes, limited differentiation has occurred despite an ancient Y-translocation of *amh* (the ligand of *amhr2*) (Pan et al., 2019). As an extreme example, it was shown that sturgeons possess an ancient female-specific region that is only 16 kb (Kuhl et al., 2021).

Although we found a 500-kb segment that is composed of tandem duplications and initially thought it would represent an expansion of the YSR, we found that CNVs in this region vary independently of sex, and hence are not causally related to sex determination (Figure S5). The vicinity to such repeats might, however, help explain the frequent rearrangements, including the translocation from chromosome 20 in the *Amphilophus–Archocentrus* ancestor and then to chromosome 6 in *Ar. centrarchus*. Recurrent rearrangements were also reported in northern pikes (Pan et al., 2019). Repeats are known to be a cause of recurrent chromosomal rearrangements by promoting nonhomologous exchange.

4.3 | The impact of long-reads on the discovery of ecologically meaningful SVs

It is becoming increasingly clear in many different biological systems that structural variations (SVs) are a major cause of phenotypic variation (Merot et al., 2020). This has already been proven for the Midas cichlid system, where the application of long-reads has resulted in the discovery of the two first causal variants in this system. Besides identifying the probable causal basis of sex determination in the current study, we used the same PacBio HiFi data to concomitantly map the probable causal genetic variant underlying the dark/gold colour polymorphism. This variant, which turned out to be a transposon insertion, had eluded us for over a decade (Kratochwil et al., 2022). That a single long-read sequencing run could have such a dramatic outcome is a testament to the power of these methods.

The level of contiguity and congruence to the Midas reference genome (Figures S2 and S3) that we obtained in our genome assembly was surprising, given that the original intention of our experiment using one single run of each PacBio HiFi and Bionano was solely to assemble the missing YSR. The diminished

heterozygosity of the advanced intercross specimen that we used is likely to have facilitated the de novo assembly. Nearly all chromosomes comprised fewer than four contigs. HiFi reads alone (or de novo assemblies thereof) are of sufficient size for population-based investigations of SV frequency and distribution that are needed to advance the field (Merlot et al., 2020). Bionano optical maps were crucial (i) to scaffold Y-haplotigs into chromosome 4 and (ii) to validate the highly complex repetitive region adjacent to the YSR, where the size of repeat units exceeded the 20-kb HiFi read length (Figure 2b). Bionano data are well suited for haplotype phasing and assembly, but current software tools do not support these methods. To circumvent this limitation, we employed a manual approach to reconstruct X and Y chromosomes, which was successful, but labour-intensive. The proprietary Bionano software tools ensure ease-of-use, convenient visualization of SVs and an overall superb functionality. However, their closed nature might hamper development as the community seems drawn towards other types of data, judging from recently published software such as HIFIASM (Cheng et al., 2021). Given the widespread interest in haplotype-aware methods, we expect this to change soon.

4.4 | Reference bias generates GWA false-positives

Filling sequence gaps in reference assemblies using long-reads is also consequential for the application of population-based GWA, as it is for the estimation of other population genetic parameters based on read mapping (Brandt et al., 2015; Chen et al., 2021). In a preliminary attempt to map sex determination, we used short-read data from Kautt et al. (2020) and the female reference genome to perform GWA analysis for sex, revealing four distinctive peaks of association. These included markers on chromosomes 4 and 20 that were under complete linkage disequilibrium. The inspection of mapping coverage showed that these were false-positives originating from distorted genotypic frequencies in males due to paralogous read-mapping. Due to the absence of the YSR, YSR-reads mapped to similar regions elsewhere in the reference such as the conserved segments of the ancestral copy of *amhr2* (chromosome 20) and TEs in other regions. Since SVs appear to be common, this observation reinforces the need of moving beyond reference genomes of single individuals towards pangenomes (Chen et al., 2021). Given the ubiquity of TEs in sex-determining regions and throughout the genome, caution must be exercised in the interpretation of GWAS results in the absence of pangenomic and genetic linkage data.

5 | CONCLUSIONS

The unprecedented ability to generate chromosome-level assemblies and map complex causal variants has already led to many exciting discoveries in both established and nonmodel systems (Kratochwil et al., 2022; Merot et al., 2020). With the prospect of applying accurate long-read sequencing to natural populations already in sight, coupling these methods with forward-genetics and ecological data will rapidly move the field towards realizing the central aim of uncovering general genetic mechanisms that govern evolutionary processes in the wild.

CONFLICT OF INTEREST

We declare we have no competing interests.

AUTHOR CONTRIBUTIONS

F.H., A.M. and C.F.K. conceived the study. F.H. and G.M.S. established the mapping panel. A.F.K. and G.M.S. performed phenotyping. F.H. and G.M.S. performed linkage mapping. A.N., A.F.K. and C.F.K. performed preliminary analysis on short read data and GWA. C.N. and F.H. collected and analysed the long-read and optical mapping data. C.N. performed all other analyses (annotation, CNV detection, comparative genomics, etc.) and wrote the first draft of the manuscript. All authors contributed to the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The ddRAD catalogue, linkage mapping input file for R/qtl, Bionano raw molecules and male genomes are available at Dryad (https://doi.org/10.5061/dryad.n02v6wx05). The PacBio reads are available at SRA (accession PRJNA694028).

ORCID

Camila L. Nacif https://orcid.org/0000-0001-7016-7242
Claudius F. Kratochwil https://orcid.org/0000-0002-5646-3114
Andreas F. Kautt https://orcid.org/0000-0001-7792-0735
Alexander Nater https://orcid.org/0000-0002-4805-5575
Gonzalo Machado-Schiaffino https://orcid.org/0000-0002-4049-3247
Axel Meyer https://orcid.org/0000-0002-0888-8193
Frederico Henning https://orcid.org/0000-0002-5359-504X

REFERENCES

Adolfi, M. C., Nakajima, R. T., Nóbrega, R. H., & Schartl, M. (2019). Intersex, hermaphroditism, and gonadal plasticity in vertebrates: Evolution of the müllerian duct and amh/amhr2 signaling. *Annual Review of Animal Biosciences, 7*, 149–172. https://doi.org/10.1146/annurev-animal-020518-114955

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology, 20*(1), 1–17. https://doi.org/10.1186/s13059-019-1829-6

Ashman, T. L., Bachtrog, D., Blackmon, H., Goldberg, E. E., Hahn, M. W., Kirkpatrick, M., & Vamosi, J. C. (2014). Tree of sex: A database of sexual systems. *Scientific Data*, 1, 1–8. https://doi.org/10.1038/sdata.2014.15

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, 12(7), 1–13. https://doi.org/10.1371/journal.pbio.1001899

Barluenga, M., Stolting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, 439(7077), 719–723. https://doi.org/10.1038/nature04325

Blaser, O., Grossen, C., Neuenschwander, S., & Perrin, N. (2013). Sex-chromosome turnovers induced by deleterious mutation load. *Evolution*, 67(3), 635–645. https://doi.org/10.1111/j.1558-5646.2012.01810.x

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, 362(6415). https://doi.org/10.1126/science.aam5979

Böhne, A., Weber, A.-A.-T., Rajkov, J., Rechsteiner, M., Riss, A., Egger, B., & Salzburger, W. (2019). Repeated evolution versus common ancestry: Sex chromosome evolution in the haplochromine cichlid *Pseudocrenilabrus philander*. *Genome Biology and Evolution*, 11(2), 439–458. https://doi.org/10.1093/gbe/evz003

Brandt, D. Y. C., Aguiar, V. R. C., Bitarello, B. D., Nunes, K., Goudet, J., & Meyer, D. (2015). Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 Genomes Project Phase I Data. *G3 Genes|Genomes|Genetics*, 5(5), 931–941. https://doi.org/10.1534/g3.114.015784

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, e4958. https://doi.org/10.7717/peerj.4958

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chalopin, D., Volff, J. N., Galiana, D., Anderson, J. L., & Schartl, M. (2015). Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Research*, 23(3), 545–560. https://doi.org/10.1007/s10577-015-9490-8

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Charlesworth, B. (1991). Evolution of sex chromosomes. *Nature*, 147(3739), 809.

Chen, N.-C., Solomon, B., Mun, T., Iyer, S., & Langmead, B. (2021). Reference flow: Reducing reference bias using multiple population genomes. *Genome Biology*, 22(1), 8. https://doi.org/10.1186/s13059-020-02229-3

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and high-performance computing Europe PMC Funders Group. *Nature Methods*, 9(8), 772. https://doi.org/10.1038/nmeth.2109.jModelTest

Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture*, 208(3–4), 191–364. https://doi.org/10.1016/S0044-8486(02)00057-1

Duan, W., Gao, F.-X., Chen, Z.-W., Gao, Y., Gui, J.-F., Zhao, Z., & Shi, Y. (2021). A sex-linked SNP mutation in amhr2 is responsible for male differentiation in obscure puffer (*Takifugu obscurus*). *Molecular Biology Reports*, 48(8), 6035–6046. https://doi.org/10.1007/s11033-021-06606-4

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology and Evolution*, 26(6), 298–306. https://doi.org/10.1016/j.tree.2011.02.008

El Taher, A., Ronco, F., Matschiner, M., & Salzburger, W. (2020). Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. *BioRxiv*. https://doi.org/10.1101/2020.10.23.335596

Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., & Guiguen, Y. (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*. *Molecular Ecology Resources*, 20(2), 531–543. https://doi.org/10.1111/1755-0998.13133

Francis, R. C., & Barlow, G. W. (1993). Social control of primary sex differentiation in the Midas cichlid. *Proceedings of the National Academy of Sciences, USA*, 90(22), 10673–10675. https://doi.org/10.1073/pnas.90.22.10673

Fraser, B. A., Whiting, J. R., Paris, J. R., Weadick, C. J., Parsons, P. J., Charlesworth, D., Bergero, R., Bemm, F., Hoffmann, M., Kottler, V. A., Liu, C., Dreyer, C., & Weigel, D. (2020). Improved reference genome uncovers novel sex-linked regions in the guppy (*Poecilia reticulata*). *Genome Biology and Evolution*, 12(10), 1789–1805. https://doi.org/10.1093/GBE/EVAA187

Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., Shu, J. J., & Mank, J. E. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biology and Evolution*, 12(6), 750–763. https://doi.org/10.1093/gbe/evaa081

Gammerdinger, W. J., & Kocher, T. D. (2018). Unusual diversity of sex chromosomes in African cichlid fishes. *Genes*, 9(10), 480. https://doi.org/10.3390/genes9100480

Gao, F.-X., Shi, Y., Duan, W., Lu, W.-J., Huang, W., Zhang, X.-J., Zhao, Z., Zhou, L. I., & Gui, J.-F. (2020). A rapid and reliable method for identifying genetic sex in obscure pufferfish (*Takifugu obscurus*). *Aquaculture*, 519(September). https://doi.org/10.1016/j.aquaculture.2019.734749

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint*, 1207(3907), 1–9. http://arxiv.org/abs/1207.3907

Godwin, J., & Roberts, R. (2018). Environmental and genetic sex determining mechanisms in fishes. In J. L. Leonard (Ed), *Transitions between sexual systems* (pp. 311–344). Springer International Publishing. https://doi.org/10.1007/978-3-319-94139-4_11

Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*. W.W. Norton.

Graves, J. A. M., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, 11(4). https://doi.org/10.1186/gb-2010-11-4-205

Hattori, R. S., Murai, Y. U., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., Fernandino, J. I., Somoza, G. M., Yokota, M., & Strüssmann, C. A. (2012). A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences, USA*, 109(8), 2955–2959. https://doi.org/10.1073/pnas.1018392109

Henning, F., Machado-Schiaffino, G., Baumgarten, L., & Meyer, A. (2017). Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. *Evolution*, 71(5), 1297–1312. https://doi.org/10.1111/evo.13206

Henning, F., & Meyer, A. (2014). The evolutionary genomics of cichlid fishes: Explosive speciation and adaptation in the postgenomic Era. *Annual Review of Genomics and Human Genetics*, 15, 417–441. https://doi.org/10.1146/annurev-genom-090413-025412

Herpin, A., Schartl, M., Depincé, A., Guiguen, Y., Bobe, J., Hua-Van, A., & Luckenbach, J. A. (2021). Allelic diversification after transposable element exaptation promoted gsdf as the master sex determining

gene of sablefish. *Genome Research*, 31(8), 1366–1380. https://doi.org/10.1101/gr.274266.120

Ieda, R., Hosoya, S., Tajima, S., Atsumi, K., Kamiya, T., Nozawa, A., Aoki, Y., Tasumi, S., Koyama, T., Nakamura, O., Suzuki, Y., & Kikuchi, K. (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(1), 1–20. https://doi.org/10.1371/journal.pone.0190635

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., & Kikuchi, K. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, 8(7), e1002798. https://doi.org/10.1371/journal.pgen.1002798

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., & Meyer, A. (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, 588(7836), 106–111. https://doi.org/10.1038/s41586-020-2845-0

Kitano, J., & Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental Biology of Fishes*, 94(3), 549–558. https://doi.org/10.1007/s10641-011-9853-8

Kratochvíl, L., Stöck, M., Rovatsos, M., Bullejos, M., Herpin, A., Jeffries, D. L., & Pokorná, M. J. (2021). Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1833), 20200097. https://doi.org/10.1098/rstb.2020.0097

Kratochwil, C. F., Kautt, A. F., Nater, A., Harer, A., Liang, Y., Henning, F., & Meyer, A. (2022). An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes. *Nature Communications*, 13(1), 296. https://doi.org/10.1038/s41467-021-27685-8

Kratochwil, C. F., Liang, Y., Gerwin, J., Woltering, J. M., Urban, S., Henning, F., & Meyer, A. (2018). Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science*, 362(6413), 457–460. https://doi.org/10.1126/science.aao6809

Kuhl, H., Guiguen, Y., Höhne, C., Kreuz, E., Du, K., Klopp, C., Lopez-Roques, C., Yebra-Pimentel, E. S., Ciorpac, M., Gessner, J., Holostenco, D., Kleiner, W., Kohlmann, K., Lamatsch, D. K., Prokopov, D., Bestin, A., Bonpunt, E., Debeuf, B., Haffray, P., … Stöck, M. (2021). A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes. *Philosophical Transactions of the Royal Society B*, 376(1832), 20200089. https://doi.org/10.1098/rstb.2020.0089

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Kuwamura, T., Sunobe, T., Sakai, Y., Kadota, T., & Sawada, K. (2020). Hermaphroditism in fishes: An annotated list of species, phylogeny, and mating system. *Ichthyological Research*, 67(3), 341–360. https://doi.org/10.1007/s10228-020-00754-6

Lässig, M., Mustonen, V., & Walczak, A. M. (2017). Predicting evolution. *Nature Ecology & Evolution*, 1(3), 1–9. https://doi.org/10.1038/s41559-017-0077

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map

format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., Nagahama, Y., Kocher, T. D., & Wang, D. (2015). 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 Genetics*, 11(11), e1005678. https://doi.org/10.1371/journal.pgen.1005678

Lichilín, N., El Taher, A., & Böhne, A. (2021). Sex-biased gene expression and recent sex chromosome turnover. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1833). https://doi.org/10.1098/rstb.2020.0107

Manchanda, N., Portwood, J. L., Woodhouse, M. R., Seetharam, A. S., Lawrence-Dill, C. J., Andorf, C. M., & Hufford, M. B. (2020). GenomeQC: A quality assessment tool for genome assemblies and gene structure annotations. *BMC Genomics*, 21(1), 193. https://doi.org/10.1186/s12864-020-6568-2

Massague, J., Attisano, L., & Wrana, J. L. (1994). The TGF-beta family and its composite receptors. *Trends in Cell Biology*, 4(5), 172–178. https://doi.org/10.1016/0962-8924(94)90202-x

Merot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology and Evolution*, 35(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., & Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, 585(7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., & Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*, 191(1), 163–170. https://doi.org/10.1534/genetics.111.137497

Nagahama, Y., Chakraborty, T., Paul-Prasanth, B., Ohta, K., & Nakamura, M. (2021). Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiological Reviews*, 101(3), 1237–1308. https://doi.org/10.1152/physrev.00044.2019

Nakamoto, M., Uchino, T., Koshimizu, E., Kuchiishi, Y., Sekiguchi, R., Wang, L., Sudo, R., Endo, M., Guiguen, Y., Schartl, M., Postlethwait, J. H., & Sakamoto, T. (2021). A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS Genetics*, 17(8), e1009705. https://doi.org/10.1371/journal.pgen.1009705

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., & Schartl, M. (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences, USA*, 99(18), 11778–11783. https://doi.org/10.1073/pnas.182314699

Nowakowska, B. A., de Leeuw, N., Ruivenkamp, C. A. L., Sikkema-Raddatz, B., Crolla, J. A., Thoelen, R., Koopmans, M., den Hollander, N., van Haeringen, A., van der Kevie-Kersemaekers, A.-M., Pfundt, R., Mieloo, H., van Essen, T., de Vries, B. B. A., Green, A., Reardon, W., Fryns, J.-P., & Vermeesch, J. R. (2012). Parental insertional balanced translocations are an important cause of apparently de novo CNVs in patients with developmental anomalies. *European Journal of Human Genetics*, 20(2), 166–170. https://doi.org/10.1038/ejhg.2011.157

Oldfield, R. G. (2009). Growth patterns in Midas Cichlids are not consistent with a hypothesis of socially controlled sex determination. *Copeia*, 1, 71–77. https://doi.org/10.1643/CG-06-115

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, 28(21), 4709–4724. https://doi.org/10.1111/mec.15245

Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.-M., Bobe, J., Concordet, J.-P., Parrinello, H., Journot, L., Klopp, C., Lluch, J., Roques, C., Postlethwait, J., Schartl, M., Herpin, A., & Guiguen, Y. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genetics*, 15(8), e1008013. https://doi.org/10.1371/journal.pgen.1008013

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, 21(1), 1–31. https://doi.org/10.1186/s13059-020-02097-x

Pennell, M. W., Mank, J. E., & Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology*, 27(19), 3950–3963. https://doi.org/10.1111/mec.14540

Qu, M., Liu, Y., Zhang, Y., Wan, S., Ravi, V., Qin, G., Jiang, H., Wang, X., Zhang, H., Zhang, B. O., Gao, Z., Huysseune, A., Zhang, Z., Zhang, H., Chen, Z., Yu, H., Wu, Y., Tang, L. U., Li, C., … Lin, Q. (2021). Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Science Advances*, 7(34), 1–12. https://doi.org/10.1126/sciadv.abg5196

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rafati, N., Chen, J., Herpin, A., Pettersson, M. E., Han, F., Feng, C., Wallerman, O., Rubin, C.-J., Péron, S., Cocco, A., Larsson, M., Trötschel, C., Poetsch, A., Korsching, K., Bönigk, W., Körschen, H. G., Berg, F., Folkvord, A., Kaupp, U. B., … Andersson, L. (2020). Reconstruction of the birth of a male sex chromosome present in Atlantic herring. *Proceedings of the National Academy of Sciences, USA*, 117(39), 24359–24368. https://doi.org/10.1073/pnas.2009925117

Recknagel, H., Elmer, K. R., & Meyer, A. (2013). A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (Amphilophus spp.) Obtained by massively parallel DNA sequencing (ddRADSeq). *G3: Genes|Genomes|Genetics*, 3(1), 65–74. https://doi.org/10.1534/g3.112.003897

Reichwald, K., Petzold, A., Koch, P., Downie, B. R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., Sahm, A., Szafranski, K., Taudien, S., Groth, M., Arisi, I., Weise, A., Bhatt, S. S., Sharma, V., Kraus, J. M., … Platzer, M. (2015). Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell*, 163(6), 1527–1538. https://doi.org/10.1016/j.cell.2015.10.071

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. https://doi.org/10.1038/nbt.1754

Schartl, M., Schmid, M., & Nanda, I. (2016). Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma*, 125(3), 553–571. https://doi.org/10.1007/s00412-015-0569-y

Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics*, 19(6), 329–346. https://doi.org/10.1038/s41576-018-0003-4

Sember, A., Nguyen, P., Perez, M. F., Altmanová, M., Ráb, P., & Cioffi, M. D. B. (2021). Multiple sex chromosomes in teleost fishes from a cytogenetic perspective: State of the art and future challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1833). https://doi.org/10.1098/rstb.2020.0098

Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin-I, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., Hamaguchi, S., Sakaizumi, M., & Naruse, K. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature Communications*, 5(1), 4157. https://doi.org/10.1038/ncomms5157

Tao, W., Conte, M. A., Wang, D., & Kocher, T. D. (2021). Network architecture and sex chromosome turnovers. *BioEssays*, 43(3), 2000161. https://doi.org/10.1002/bies.202000161

Tao, W., Xu, L., Zhao, L., Zhu, Z., Wu, X., Min, Q., Wang, D., & Zhou, Q. I. (2021). High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes. *Molecular Ecology Resources*, 21(2), 543–560. https://doi.org/10.1111/1755-0998.13273

Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25), 731. https://doi.org/10.21105/joss.00731

Vicoso, B. (2019). Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology and Evolution*, 3(12), 1632–1641. https://doi.org/10.1038/s41559-019-1050-8

Xiong, P., Hulsey, C. D., Fruciano, C., Wong, W. Y., Nater, A., Kautt, A. F., Simakov, O., Pippel, M., Kuraku, S., Meyer, A., & Franchini, P. (2021). The comparative genomic landscape of adaptive radiation in crater lake cichlid fishes. *Molecular Ecology*, 30(4), 955–972. https://doi.org/10.1111/mec.15774

Xue, L., Gao, Y. U., Wu, M., Tian, T., Fan, H., Huang, Y., Huang, Z., Li, D., & Xu, L. (2021). Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. *Genome Biology*, 22(1), 1–20. https://doi.org/10.1186/s13059-021-02430-y

Yang, L., Xu, Z., Zeng, H., Sun, N., Wu, B., Wang, C., Bo, J., Li, L., Dong, Y., & He, S. (2020). FishDB: an integrated functional genomics database for fishes. *BMC Genomics*, 21(1), 1–5. https://doi.org/10.1186/s12864-020-07159-9

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.