# Identification of the male-specific region on the guppy Y Chromosome from a haplotype-resolved assembly

Kang Du,[1] Oliver Deusch,[2] Ilja Bezrukov,[2] Christa Lanz,[2] Yann Guiguen,[3] Margarete Hoffmann,[2] Anette Habring,[2] Detlef Weigel,[2] Manfred Schartl,[1,4,5,6] and Christine Dreyer[2,6]

[1]*Xiphophorus Genetic Stock Center, Institute for Molecular Life Sciences, Texas State University, San Marcos, Texas 78666, USA;* [2]*Max Planck Institute for Biology Tübingen, Department of Molecular Biology, 72076 Tübingen, Germany;* [3]*INRAE, LPGP, 35073 Rennes, France;* [4]*Theodor Boveri Institute, Developmental Biochemistry, Biocenter, University of Würzburg, 97074 Würzburg, Germany;* [5]*Research Department for Limnology, University of Innsbruck, 5130 Mondsee, Austria*

**The guppy Y Chromosome has been a paradigmatic model for studying the genetics of sex-linked traits and Y Chromosome–driven evolution for more than a century. Despite strong efforts, knowledge on genomic organization and molecular differentiation of the sex chromosome pair remains unsatisfactory and partly contradictory with respect to regions of reduced recombination. Especially the border between pseudoautosomal and male-specific regions of the Y has not been defined so far. To circumvent the problems in assigning the repeat-rich differentiated hemizygous or heterozygous sequences of the sex chromosome pair, we sequenced a YY male generated by a cross of a sex-reversed Maculatus strain XY female to a normal XY male from the inbred Guanapo population. High-molecular-weight genomic DNA from the YY male was sequenced on the Pacific Biosciences platform, and both Y haplotypes were reconstructed by Trio binning. By mapping of male specific SNPs and RADseq sequences, we identify a single male specific-region of ~5 Mb length at the distal end of the Y (MSY). Sequence divergence between X and Y in the segment is on average five times higher than in the proximal part in agreement with reduced recombination. The MSY is enriched for repeats and transposons but does not differ in the content of coding genes from the X, indicating that genic degeneration has not progressed to a measurable degree.**

**[Supplemental material is available for this article.]**

Sex chromosomes evolve differently from the rest of the genome owing to reduced recombination. This process generates on Y Chromosomes a male-specific region (MSY) that is diverging in sequence and structural organization from the homologous region on the X. Because of reduced recombination, genes within the MSY can become restricted to the male sex (Khoo et al. 1999; Lindholm and Breden 2002). The guppy was the first vertebrate in which sex linkage of phenotypic traits was described. In this fish, males develop highly variable, yet heritable nuptial patterns. Genes that control the characteristic color traits could be genetically mapped on the X and Y Chromosomes (Winge 1922; Khoo et al. 1999; Lindholm and Breden 2002; Tripathi et al. 2009a). A genetic map based on EST- and BAC-derived markers was generated, and LG12 was identified as the sex chromosome (Tripathi et al. 2009a,b). Several markers including the most distal genetic marker M_229 (located in the cyclin I gene) (Tripathi et al. 2009a,b) differentiating X and Y were located by molecular cytogenetic analysis on the homologous chromosomes of different guppy populations and the two sister species *Poecilia wingei* and *Poecilia obscura* (Nanda et al. 2014). This analysis revealed polymorphisms in heterochromatin content of the Y as well as differences in distance of the genetic marker M_229 to the physical chromosome end between populations. The sex determination locus (SDL) was mapped to the most distal region of the Y (Winge 1922; Tripathi et al. 2009a,b), but its molecular identity has remained unknown. The ornamental male coloration is a combination of autosomal and many sex chromosome–linked genes (Winge 1927; Kirpichnikov 1981; Lindholm and Breden 2002; Paris et al. 2022; van der Bijl et al. 2024). Similarly, we still lack promising candidates for genes controlling the Y-linked pigmentation traits. In contrast to the MSY, the pseudoautosomal region (PAR) is found on both sex chromosomes and is considered as freely recombining between the X and Y Chromosomes. Yet, the literature contains conflicting results regarding the evolution of the different regions of the sex chromosomes as well as the extent of recombination suppression on the Y (Wright et al. 2017; Bergero and Charlesworth 2019; Darolti et al. 2019; Kirkpatrick et al. 2022).

Two whole-genome assemblies of the common (or Orinoco) guppy (*Poecilia reticulata*) from the Guanapo river in West Trinidad have been published previously: a female genome assembled from Illumina short reads (Künstner et al. 2016) and a male genome assembled from Pacific Biosciences (PacBio) Continuous Long Reads (CLR) (Fraser et al. 2020). A genetic map derived from crosses between individuals from Cumaná (Venezuela) and upper Quare river served to scaffold the assemblies to chromosome level (Tripathi et al. 2009a,b).

[6]**These authors contributed equally to this work.**
**Corresponding authors: phch1@biozentrum.uni-wuerzburg.de, christine.dreyer@tuebingen.mpg.de**
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.279582.124. Freely available online through the *Genome Research* Open Access option.

© 2025 Du et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

**Figure 1.** Adult male guppy phenotypes. (*A*) Male of the Maculatus strain, genotype $X_{mac}Y_{mac}$. (*B*) Male of the Guanapo strain, genotype $X_{Gua}Y_{Gua}$. (*C*) $F_1$ male with sex chromosome constitution $Y_{gua}Y_{mac}$. Note that both the characteristic Maculatus black spot on the dorsal fin (arrowhead) and Guanapo black spot on the anterior trunk (arrow), as well as the Guanapo tail pattern (star), are expressed in the $Y_{gua}Y_{mac}$ male.

Unfortunately, the male guppy genome (Fraser et al. 2020) does not distinguish X and Y Chromosome (Chr) sequences and the maternal and paternal haplotypes of Chr 12 are collapsed into a single scaffold. To obtain a haplotype-resolved male genome, we made use of the fact that male-to-female sex-reversed XY individuals occur spontaneously in the Maculatus strain, likely caused by an as-yet-unidentified autosomal factor. Crosses to regular XY males can produce YY offspring, but these die as embryos, likely owing to a recessive lethal-effect mutation on the Maculatus Y. In XY males, this mutation is complemented by a gene on the X. The lethal-effect mutation on the Maculatus Y can also be complemented by alleles on Y Chromosomes from other populations (Winge 1922, 1938). Thus, to produce an adult YY male for genome sequencing, we crossed an XY female from the Maculatus strain to an XY male of the Guanapo strain.

Trio binning is a powerful method to sort PacBio sequencing long reads according to their maternal and paternal origin (Koren et al. 2018). Thus, we aimed at sequencing and assembling a chromosome-level genome of a YY male from an inter-populational cross using trio binning to resolve both Y Chromosome haplotypes. This genomic resource was expected to inform about the structure and gene content of the guppy Y Chromosome, in particular about the localization and extent of PARs, segments of reduced recombination, and male-specific regions.

## Results

### DNA sequencing and trio binning

To obtain a Y haplotype-resolved assembly, we crossed a Guanapo male with a sex-reversed Maculatus XY female to obtain a YY $F_1$ male offspring (Fig. 1A–C). This fish carried one Y Chromosome from the Guanapo population and the other one from the Maculatus population. The reconstructed Y haplotypes are designated $Y_{mac}$ and $Y_{gua}$.

Continuous long-read (CLR) sequencing on the PacBio Sequel II platform of high-molecular-weight DNA from this $Y_{mac}Y_{gua}$ male resulted in 170 Gb sequence, corresponding to approximately 226-fold genome coverage, assuming a genome size of ~750 Mb (Table 1A). Because DNA samples from the actual parents were not available, we generated Illumina short-read data from another Maculatus XY female and from the genomic DNA that was used for the previously published Guanapo XY male genome (Table 1B; Fraser et al. 2020). Trio binning assigned 85.9 Gb to the Guanapo haplotype and 84.0 Gb to the Maculatus haplotype. Only 0.01 Gb could not be assigned to either haplotype, and 0.1 Gb was too short to be binned (Table 1A). BUSCO completeness was 98.0% for Guanapo and 97.9% for Maculatus, with 2.6% apparently duplicated in Guanapo and Maculatus, respectively (Table 1C).

### Reconstruction of all autosomes from phased haplotype contigs

Trio binning allowed reconstructing all 23 chromosomes as phased haplotypes from the Guanapo and Maculatus strains, using the published guppy autosome sequences from the Guanapo XY individual as reference for scaffolding. We then aligned the resulting Guanapo and Maculatus autosomes to their orthologous *Xiphophorus hellerii* and *Xiphophorus maculatus* chromosomes as

**Table 1.** DNA sequencing overview

**A. PacBio Sequel II (CLR)**

| | Total bases (Gb) | No. of sequences | Max length (bp) | N50 (bp) | N90 (bp) |
|---|---|---|---|---|---|
| All reads | 170.06 | 8,503,777 | 521,561 | 29,768 | 11,540 |
| Guanapo | 85.91 | 4,162,756 | 397,496 | 29,755 | 11,719 |
| Maculatus | 84.01 | 4,094,477 | 521,561 | 29,814 | 11,452 |
| Unclassified | 0.01 | 7328 | 22,512 | 1688 | 1097 |

**B. Illumina (2 × 150 bp, PCR-free)**

| | | |
|---|---|---|
| Maculatus XY female | 146 M cluster | 44 Gb |
| Guanapo XY male | 146 M cluster | 43 Gb |

**C. Assembly statistics[a]**

| | Total bases (Mb) | No. of sequences | Max length (bp) | N50 | L50 | BUSCO completeness | BUSCO duplicated |
|---|---|---|---|---|---|---|---|
| Guanapo | 745.3 | 23 | 46,441,515 | 34,623,888 | 11 | 98.00% | 2.60% |
| Maculatus | 745.2 | 23 | 46,441,515 | 34,623,888 | 11 | 97.90% | 2.60% |

[a]These final genomes were assembled using the same reference genome as guide, resulted in highly similar statistics.

the most closely related species from which high-quality chromosome-level assemblies are available (Supplemental Fig. S1). These alignments showed, in general, an overall conservation of synteny of guppy autosomes to their *Xiphophorus* homologs (Lu et al. 2023). It also revealed that most of the inversions that distinguish the guppy and *X. maculatus* coincide in length and approximate position with inversions that were also observed between *X. maculatus* and *X. hellerii* (Supplemental Fig. 1 in Lu et al. 2023). As *X. maculatus* is a derived species and *X. hellerii* is more basal in the phylogenomic trees (Lu et al. 2023; Du et al. 2024), these inversions must be caused by events in the *X. maculatus* lineage. Confirming previous assemblies, the guppy female linkage group 2 combines medaka Chr 2 and 21 (Künstner et al. 2016). Chr 2 is a fusion of Chr 7 with Chr 24 (Fraser et al. 2020) of *Xiphophorus*, which represents the basal poeciliid karyotype (Cimino 1974). This fusion was also observed in the genome assembly of *Poecilia picta* (Metzger et al. 2021).

## Assembly, genomic organization, and gene content of the Y Chromosomes

As guidance for assembly of Chr 12, we used a modified version of the published Guanapo male Chr 12 (LR880656.1) (Fraser et al. 2020). Revisiting previously published Hi-C data (Fraser et al. 2020), we changed the order of the XY genome contigs (Supplemental Table S1). We used 11 Guanapo and nine Maculatus contigs >200 kb for a primary assembly of $Y_{gua}$ (28.3 Mb) and $Y_{mac}$ (27.6 Mb) (Supplemental Table S2; for details, see Supplemental Material).

To investigate the gene content for each Y haplotype, we conducted a genome annotation with an in-house pipeline adapted from a previous study (Du et al. 2022), which synthesizes evidence from homology aligning, transcriptome mapping, and ab initio gene prediction. On the $Y_{gua}$ 1159 protein coding genes were annotated, with 932 (80.4%) showing transcriptome support and 1123 (96.9%) with BLAST hits in RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) or Swiss-Prot database (https://www.uniprot.org/). On the $Y_{mac}$, 1122 protein coding genes were annotated, of which 957 (85.3%) were transcriptome supported and 1090 (97.1%) had BLAST hits in RefSeq or Swiss-Prot (Table 2). Although there is congruence in general, discrepancies between both haplotypes are seen in which genes of one haplotype are on short unplaced contigs in the other haplotype. On the other hand, binning errors may have caused apparent gene duplications on $Y_{gua}$.

## Comparison of Y haplotypes to published female and male guppy Chr 12 sequences

The reconstructed $Y_{gua}$ and $Y_{mac}$ haplotypes were aligned to each other and to the published chimeric X/Y male Chr 12 (LR880656.1) (Fraser et al. 2020), as well as to LG12 (NC_024342.1) from the XX female assembly (Fig. 2; Supplemental Fig. S2; Künstner et al. 2016). Alignment gaps between $Y_{gua}$ and LR880656.1 are caused by several scaffolds that had remained unplaced in the previous assembly (000013F, 000200F, 000250F, 0000181F, and 000149F) (Supplemental Table S1; Fraser et al. 2020; Whiting et al. 2022).

A genetic linkage map based on SNP markers deduced from EST and BAC end sequences (Tripathi et al. 2009b) had served as guide for the assembly of both the previously published female and male genomes. The most distal sex chromosomal marker with consistently heterozygous SNPs in males but not females is M_229 (cyclin I, g876 on Ygua, g873 on Ymac) (Tripathi et al. 2009a,b; Lisachov et al. 2015; Bergero et al. 2019). Physical location of cyclin I and several other genetic markers has been confirmed cytologically by chromosome FISH with BACs containing these marker genes. This revealed consistent X and Y Chromosome organization in guppies of widely diverging geographic origin (Nanda et al. 2014).

Alignments of *X. maculatus* Chr 8 (Lu et al. 2023), which is homologous to guppy Chr 12, to the $Y_{gua}$ and $Y_{mac}$ haplotypes revealed a long inversion of ~3.8 Mb, as also seen with LR880656.1 (Supplemental Fig. S8 in Fraser et al. 2020). In both haplotypes, this inversion is found in the middle of >9-Mb-long contigs, Gua_tig00000294, and Mac_tig00000028, respectively. An inversion of similar length and position also distinguishes *X. maculatus* and *X. hellerii* Chr 8 (Lu et al. 2023). When $Y_{Gua}$ was aligned to *X. hellerii* Chr 8, it turned out that the inversion is specific to *X. maculatus*, suggesting that guppy Chr 12 is more similar to *X. hellerii* Chr 8, which represents the ground state for *Xiphophorus* (Fig. 3). To validate our inferences, we scaffolded the female Chr 12 from existing female contigs (Künstner et al. 2016) using the $Y_{gua}$ and $Y_{mac}$ haplotypes as guides. This resulted in a reconstructed Guanapo X Chromosome ($X_{gua}$) and a virtual Maculatus X

**Table 2.** Annotation statistics of $Y_{gua}$ and $Y_{mac}$ MSY and pseudoautosomal region (PAR)

| | MSY (23.5 Mb-end) | | PAR (1–23.5 Mb) | |
|---|---|---|---|---|
| | $Y_{gua}$ | $Y_{mac}$ | $Y_{gua}$ | $Y_{mac}$ |
| Total genes | 210 | 179 | 949 | 943 |
| Single exon | 16 (7.6%) | 14 (7.8%) | 60 (6.3%) | 57 (6.0%) |
| Multiexon | 194 (92.4%) | 165 (92.2%) | 889 (93.7%) | 886 (94.0%) |
| Map to Pfam | 161 (76.7%) | 140 (78.2%) | 770 (81.1%) | 784 (83.1%) |
| Map to RefSeq/Swiss-Prot | 199 (94.8%) | 171 (95.5%) | 924 (97.4%) | 919 (97.5%) |
| With RNA support | 57 (27.1%) | 24 (13.4%) | 108 (11.4%) | 84 (8.9%) |
| Without start codon | 35 (16.7%) | 15 (8.4%) | 50 (5.3%) | 50 (5.3%) |
| Pseudo with RNA | 16 (7.6%) | 18 (10.1%) | 64 (6.7%) | 60 (6.4%) |
| Pseudo without RNA | 19 (9.0%) | 11 (6.1%) | 39 (4.1%) | 33 (3.5%) |
| Repeats | 55.85% | 54.17% | 34.53% | 34.48% |

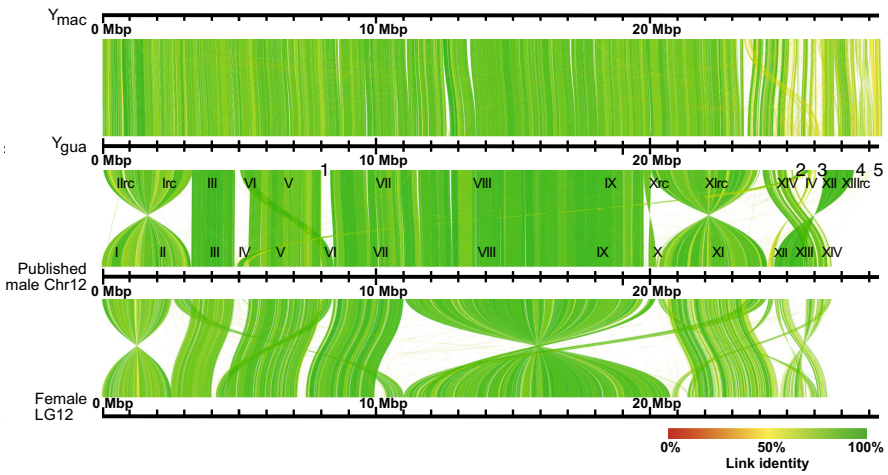


**Figure 2.** Alignment of $Y_{gua}$ and $Y_{mac}$ to published XY male Chromosome 12 (LR880656) and female LG12 (NC024342). Roman numbers specify positions of XY contigs on LR880656; for details see Supplemental Table S1. Positions of additional unplaced scaffolds of the XY assembly are indicated under $Y_{gua}$: 1, 00013F_0; 2, 000200F_0; 3, 000250F_0; 4, 000181F_0; and 5, 000149F_0.

Chromosome ($X_{mac}$). Alignments (Fig. 4) revealed that the large inversion previously detected between the two Y haplotypes and the published LG12 (Fig. 2; Künstner et al. 2016) was almost certainly caused by a previous scaffolding error. This inversion was also noted in the assembly of the male guppy genome (Fraser et al. 2020) and by comparison to the assembly of *P. picta* (Metzger et al. 2021). Total chromosome lengths differ by ~4.2 Mb between the Guanapo X and Y reconstructions. This difference is mainly caused by Y-sequence regions lacking counterparts on X, especially near the chromosome ends (Supplemental Figs. S3, S4).

## Identification of the male-specific region (MSY) on both Y Chromosomes

In another attempt to identify the MSY on $Y_{gua}$ and $Y_{mac}$, we first aligned both to the reassembled version of the female Chr 12, which showed that both Y Chromosomes ($Y_{gua}$ and $Y_{mac}$) diverged from the X Chromosome at the distal end with a sharp transition at ~23.5–24 Mb (Fig. 5A) The average divergence in the distal end is 6% sequence difference (SNPs and indels), around fivefold higher than in the proximal part (average, 1%). Although the higher divergence of the distal part is consistent with lower effective recombination, in agreement with known features of the MSY, our data do not support that there is a second region between 15 and 21 Mb, where X and Y do not recombine ("stratum 2"), as has been postulated previously (Wright et al. 2019), nor that there is an even more proximal nonrecombining region ("MSNR1"), as inferred from meiotic mapping (Tripathi et al. 2009a). The genomic structure of the guppy sex chromosomes with a 4 Mb large region of heterozygosity considerably higher compared with autosomes and the proximal 14 Mb of LG 12 showing autosomal sequence divergence levels is also not in line with previous studies on recombination rates (Bergero et al. 2019) that proposed a single small (1–2 Mb) PAR at the tip of the Y Chromosome. These discrepancies may be explained by the usage of other populations and even different guppy species as parents for mapping crosses (*P. obscura*, *P. wingei*).

To further confirm the location of the MSY, male-specific RAD-tags were generated from four different guppy strains (including three populations from East and West Trinidad and from Venezuela) (Supplemental Data Set 1) and mapped to the assembly (Fig. 5B, red bars). These specific RAD-tags are enriched in the distal region from 20 Mb to the very end at 28.5 Mb. Adding RAD-tags from a separate analysis of guppies from the Caroni swamp in Trinidad reduced the MSY candidate region further to the 23.5 to 28.5 Mb interval (Fig. 5B, blue bars). The different proximal

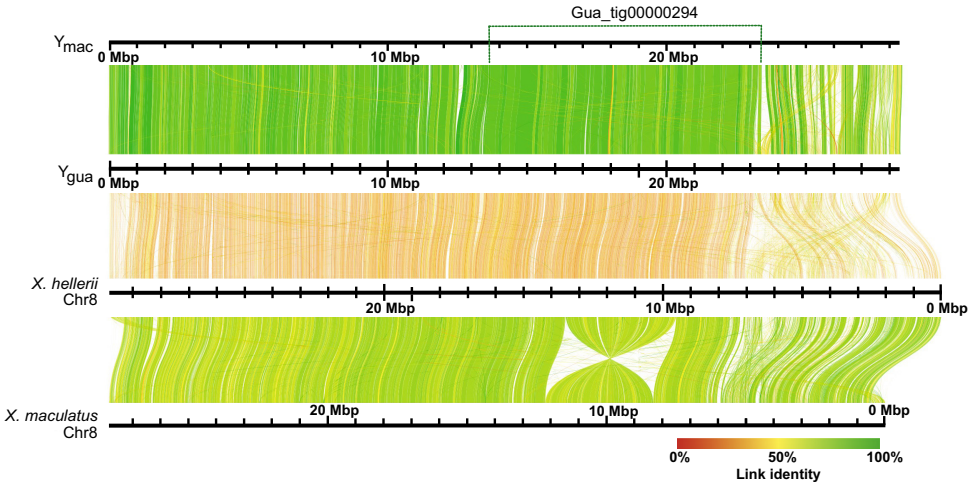


**Figure 3.** Alignment of $Y_{gua}$ and $Y_{mac}$ haplotypes to *Xiphophorus* Chromosome 8. A 3.5 Mb inversion distinguishes *X. hellerii* and *X. maculatus*. The position of the long contig Gua_tig00000294 that spans the inversion of *X. maculatus* Chr 8 is indicated at the *top*. The position of a short proximal sequence of the guppy Y also conforms to *X. hellerii* Chr 8, whereas it is translocated to the chromosome end in *X. maculatus*. $Y_{gua}$ and $Y_{mac}$ conform to one another and to *X. hellerii* Chr 8, with most divergence seen from 23 Mb to the end of guppy Y haplotypes.

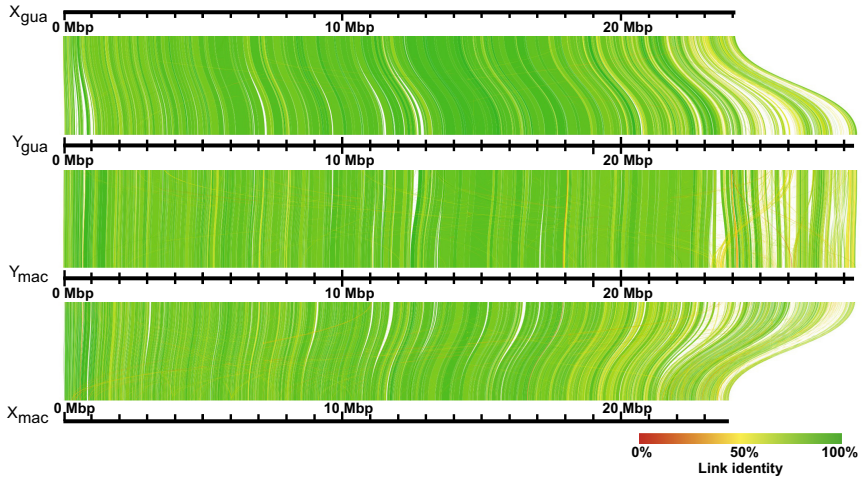


**Figure 4.** Alignment of reconstructed X Chromosomes to $Y_{gua}$ and $Y_{mac}$ haplotypes. $Y_{gua}$ and $Y_{mac}$ assembled as described in Supplemental Table S2 are aligned to $X_{gua}$ and a virtual $X_{mac}$. Both X Chromosomes were rescaffolded from the female Guanapo contigs using $Y_{gua}$ and $Y_{mac}$ as templates.

borders of the MSY delineated by the two RAD-tag data sets could indicate that suppression of recombination expanded differently in different populations, a phenomenon noticed frequently in other species (McKinney et al. 2020; Hallast et al. 2023). In a different approach using Pool-seq data for calling male-specific SNPs from a related species, *P. obscura* (Supplemental Data Set 2), also the 24 to 25 Mb segment has a higher SNP density than the rest of the chromosome (Supplemental Fig. S5). The most distal region from 26 to 28.5 Mb exhibits a much lower density of male-specific SNPs than the rest of the chromosome. The absence of SNPs at the very end may be caused by alignment failures between the reference genome and the Pool-seq reads owing to a high divergence of Y Chromosomes between the two species. Length polymorphism between the distal regions of X and Y and between Y Chromosomes of different populations has been documented by fluorescence in situ hybridization and C-banding (Nanda et al. 1990, 2014). However, the region between 24 and 25 Mb appears to be a conserved MSY between *P. reticulata* and *P. obscura*.

Studies of sex chromosomes in many species have revealed that, owing to reduced recombination, transposable elements accumulate in the MSY (Chalopin et al. 2015; Štundlová et al. 2022). Both guppy Y Chromosomes have an increased repeat content in the terminal region (22–28 Mb) (Supplemental Table S3), with these elements being significantly younger than those in the proximal part (Supplemental Fig. S6). Such sharp borders were also seen in sex chromosomes of seahorse (Long et al. 2023) and splitfin (Du et al. 2022) without any inversion identified. Accumulation of helitron-4 elements occurred on different locations in the MSY of $Y_{gua}$ and $Y_{mac}$, indicating independent evolution of the Y Chromosome in different populations of the guppy.

The MSY contains several genes that are known to be related to sperm structure and function and the gene encoding sepiapterin reductase (*spra*), which is a key enzyme for the production of the red and yellow pigment of erythrophores and xanthophores (Table 3; Braasch et al. 2007; Schartl et al. 2016).

To further characterize the guppy MSY, we reanalyzed a transcriptome data set from male and female adult tissues (Sharma et al. 2014) and identified on the entire $Y_{gua}$ Chromosome 45 genes with significant male expression bias in the gonads. Of these, 33 are also male biased in brain and tail tissue. Male biased genes in the distal region of the Y Chromosomes include three genes associated with spermatogenesis, namely, *sperm associated antigen 8*, *flagellar Pf20-like*, and *cilia- and flagella-associated protein 53* (Supplemental Table S4; for a comparison of all genes annotated on the $Y_{gua}$ and $Y_{mac}$ haplotypes, see Supplemental Table S5). Genetic mapping predicted the putative SDL of the guppy to be located distal of the *cyclin I* gene (Tripathi et al. 2009b), which is located at 25.78 Mb of $Y_{gua}$. Therefore, the SDL might be located between *cyclin I* and 27 Mb, provided that the information from the crosses between Cumaná and Upper Quare guppies (probably laboratory-reared *P. obscura*) applies also to Guanapo guppies.

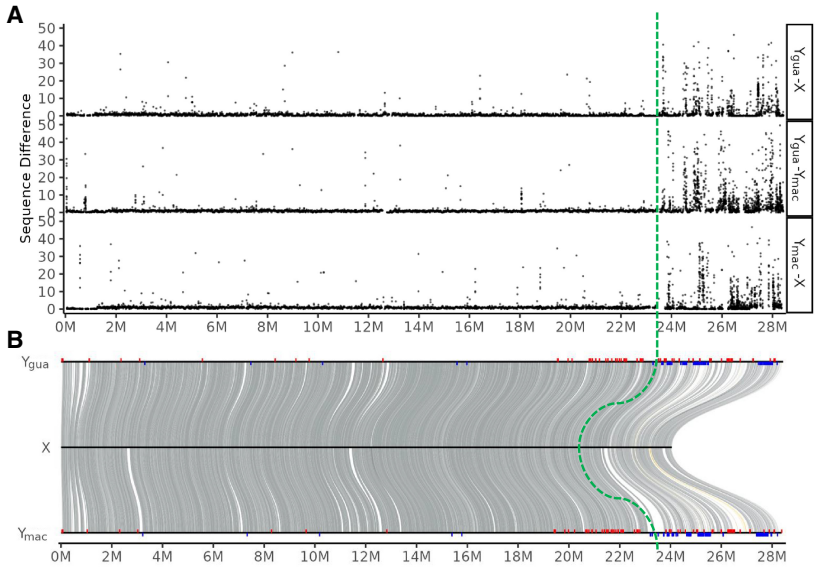


**Figure 5.** Sequence divergence between Y haplotypes and the reconstructed $X_{gua}$. (*A*) Sequence difference calculated as the number of SNPs and indels in 10 kb sliding windows along the Y Chromosomes. (*B*) $Y_{gua}$ and $Y_{mac}$ are aligned to the X. Male-specific RAD-tags from Caroni swamp guppies (blue bars) and from four different guppy strains (red bars) were mapped on the aligned $Y_{gua}$ (*top*) and $Y_{mac}$ (*bottom*) sequences. The four strains represent a broader range of geographic origins, from East Trinidad to West Venezuela. About 70% of these RAD-tags (red bars) had best hits (*e*-40) within the distal 16% of the Y (>24 Mb). Both Y haplotypes are aligned to the reconstructed $X_{gua}$ (*middle*). The green dotted line indicates the congruent MSY identified by sequence divergence and RAD-tags from the Trinidad populations of guppy.

**Table 3. Genes with functions related to male gonadal development and pigmentation in the MSY region**

| $Y_{mac}$ start | $Y_{mac}$ end | $Y_{gua}$ start | $Y_{gua}$ end | Gene annotation | Comments on function and expression |
|---|---|---|---|---|---|
| 23.793.174 | 23.796.136 | 23.577.725 | 23.580.685 | Sperm-associated antigen 8-like | Cell cycle regulation during spermiogenesis |
| 24.828.972 | 24.829.886 | 24.651.601 | 24.652.515 | Growth arrest and DNA damage-inducible protein GADD45 gamma | Regulates SRY; sex reversal candidate |
| 24.833.872 | 24.834.948 | 24.656.498 | 24.657.585 | Growth arrest and DNA damage-inducible protein GADD45 gamma | Growth arrest at puberty; male-specific trait |
| 24.940.636 | 24.943.646 | 24.771.398 | 24.774.385 | Prostate androgen-regulated mucin-like protein 1 (PARM) | Target of androgen |
| 25.311.394 | 25.334.880 | 25.141.772 | 25.165.256 | Cilia- and flagella-associated protein 44 | Sperm motility |
| 25.856.670 | 25.863.321 | 25.773.319 | 25.780.388 | Cyclin-I isoform X1 [*Poecilia reticulata*] | Last genetic marker 229; male-specific SNPs |
| 25.979.809 | 25.982.063 | 25.987.747 | 25.993.537 | Spindlin-1-like [*Xiphophorus maculatus*] | Ssty domain; spermiogenesis |
| 26.568.289 | 26.572.876 | 26.574.475 | 26.579.107 | Flagellar WD repeat-containing protein Pf20-like | Sperm motility |
| 27.005.425 | 27.007.897 | 27.199.913 | 27.202.448 | Sepiapterin reductase | Red and yellow pigment |
| 27.061.485 | 27.081.388 | 27.260.493 | 27.281.378 | Outer dense fiber protein 2 | Sperm motility |
| 27.302.394 | 27.309.962 | 27.453.812 | 27.462.891 | LMBR1 domain-containing protein 2 | Male bias expression in gonad |
| 27.329.422 | 27.332.468 | 27.484.292 | 27.487.357 | Cilia- and flagella-associated protein 53 [*P. reticulata*] | Sperm mobility; male bias |
| 28.069.253 | 28.075.419 | 28.176.114 | 28.182.692 | Hydroxysteroid dehydrogenase-like protein 2 isoform X1 | Steroid metabolism |

Positions (in base pairs) are listed for Guanapo and Maculatus haplotypes.
Comments on function and expression are based on an analysis of gene annotations. For expression patterns, see Supplemental Table S4.

In search of Y-specific genes, we then generated a list of nine genes that occur on both Y Chromosomes between 22.5 and 27 Mb but seem to be missing or are defective on the X Chromosome (Supplemental Table S6). This includes genes in ~3 Mb of sequence proximal of *cyclin I* that is rich in sex-specific RAD markers and male SNPs (Fig. 5B). When we attempted to validate by PCR and screening female raw reads and female Pool-seq reads the candidate genes differentiating the X and Y Chromosomes, we could generally not confirm that they were missing from the X Chromosome (Supplemental Tables S6, S7). Also, manual curation of the MSY assembly and exon annotation revealed no difference in the content of coding genes compared to the homologous region of the X.

## Discussion

We made use of the long-known observation that sex-reversed XY females occur spontaneously in the Maculatus strain of the guppy to produce viable males with two different Y Chromosomes. This enabled us to assemble the both the $Y_{gua}$ and $Y_{mac}$ from long-range sequence data using the powerful technique of Trio binning. The conserved MSY could be narrowed down to a few megabases of sequence length. The divergence pattern of the sex chromosome pair clearly identified only a single stratum on the guppy Y. Despite a considerable molecular differentiation of the MSY to the corresponding region of the X, our analysis revealed no structural or functional gain of annotated coding genes on the MSY, nor did we find evidence for genes found in the homologous region of the X being lost on Y. Obviously, genic degeneration has not proceeded to a noticeable degree, and no Y-specific coding gene that would act as a master male determining factor is evident from the assembly of the MSY. It cannot be excluded that such a gene failed to be assembled or missed in the annotation. On the other hand, allelic variation on the coding sequence level or in noncoding, regulatory regions of a gene present on the X and the Y could be responsible for a sex determining function (e.g., see Tang et al. 2022). Allelic variation is a common origin of sex determination (SD) genes in fish (Kitano et al. 2024). To detect such minor differences in the X and Y sequences may require the combination of more sequencing at highest possible fidelity of more individuals from different populations and different guppy species to distinguish private SNPs and sequence variation from those that are conserved across all individuals that share the same SD gene and genes encoding sex-linked traits. One possible strategy could be amplicon sequencing targeting X and Y Chromosome ends (Kersten et al. 2023). Crossing males from other populations to Maculatus X $Y_{mac}$ females to produce a range of $Y_{mac}Y_x$ males would allow comparison of different Y Chromosomes to the $Y_{gua}$ described here. Sequence comparisons between the MSY of the $Y_{gua}$ and the corresponding regions of other Y Chromosomes could also reveal the most likely candidate genes for Y-linked pigmentation patterns and other traits shared by different populations.

An important role in sex chromosome evolution is assigned to genes that are beneficial for one and/or detrimental to the other sex (Kirkpatrick and Guerrero 2014). We identified several spermatogenesis and pigmentation genes on the Y Chromosome. Some are located in the MSY (Table 3), whereas others reside in the PAR (e.g., spermatogenesis genes *morn3*, *strbp*, pigmentation genes *slc45a2* [*aim1*], *skiv2l2*, *slc31a1*). The genes in the MSY are also present in the corresponding region of the X. For the pigmentation genes (e.g., *spra*), we cannot say how they are related to male ornaments, because many male colors are encoded on the X and the PAR of the Y (Kirpichnikov 1981). Their expression, however, is dependent on a high testosterone level (Nanda et al. 2022). Future studies are necessary to reveal whether the X and Y alleles of these pigmentation candidate genes are divergent and under androgen control.

In the common guppy *P. reticulata*, the sex chromosome pair has been assigned to linkage group 12 (Künstner et al. 2016; Dor et al. 2019), and its homology with the sex chromosomes of its sister species within the subgenus *Acantophacelus*, namely, *P. obscura* and *P. wingei*, has been shown (Nanda et al. 2014; Morris et al. 2018). Also in the distantly related species *P. picta* and *Poecilia parae* (subgenus *Micropoecilia*), the Y Chromosome is derived from linkage group 12 (Bergero et al. 2019; Charlesworth et al. 2020; Metzger et al. 2021). However, in these two taxa, in strong contrast to *P. reticula*, the Y Chromosome is highly degenerated at the molecular and morphological level (Darolti et al. 2019; Charlesworth et al. 2021a,b; Nanda et al. 2022). Two opposing explanations have been put forward: either a common origin but different degrees of recombination suppression and, consequently, speed of degeneration (Darolti et al. 2019, Fong et al. 2023) or convergent evolution of the same ancestral linkage group to independently become the sex chromosome (Charlesworth et al. 2021a; Kirkpatrick et al. 2022). In this scenario, Y's have independently evolved from the same linkage group and at different evolutionary times in the lineages of *P. reticulata*/*P. obscura*/*P. wingei* on the one hand and *P. picta*/*P. parae* on the other hand. The availability of a high-quality reference genome of the common guppy with a fully assembled Y can foster population genomic studies for investigating the evolution of Y Chromosomes in this unique model species and should help to resolve the discrepancies from previous studies on sex chromosome structure and their evolutionary origin (Bergero et al. 2019; Darolti et al. 2019; Kirkpatrick et al. 2022).

## Methods

### Fish sampling and aquaculture

We made use of the fact that in the *P. reticulata* Maculatus laboratory strain (WLC 1250), spontaneous XY females arise spontaneously, which can be diagnosed by the Y-encoded black spot in the dorsal fin. Such a sex-reversed XY female was crossed to a XY male of the *P. reticulata* Guanapo strain (WLC 5856). In the resulting $F_1$ generation, the YY males were identified by their coexpression of the distinct Maculatus and Guanapo pigmentation patterns. One of these males (ID 5944-1) was used for sequencing. These fish were raised at the fish facilities of the Biocenter of the University of Würzburg following approved experimental protocols through an authorization (568/300-1870/13) of the Veterinary Office of the District Government of Lower Franconia, Germany, in accordance with the German Animal Protection Law (TierSchG).

### DNA sequencing

For long-read sequencing, high-molecular-weight genomic DNA prepared from a $Y_{gua}$ $Y_{mac}$ male was analyzed by pulse field gel electrophoresis, which revealed a main peak ~165 kb. After gentle shearing using an Eppendorf blue tip, the DNA was purified and concentrated using Ampure beads. After size-fractionation on a BluePippin Pulsed-Field gel, a fraction of 30 to 80 kb (average, 43.5 kb) was selected for construction of a PacBio CLR library following the PacBio manual. The library was sequenced on a PacBio Sequel II instrument resulting in 170 Gb sequence. This corresponds to 226-fold coverage if a genome size of 750 Mb is assumed (Table 1).

PCR-free libraries prepared from a Maculatus XY female and a Guanapo XY male with an ~300 bp insert length were sequenced with 150 bp paired end reads on an Illumina HiSeq 2000 instrument.

### Read filtering and quality trimming

Illumina short reads were trimmed using Trimmomatic (version 0.36) (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 SLIDINGWINDOW:6:30 MINLEN:60. Compared with less stringent trimming parameters, these settings reduced the amount of apparent copy number variation between the Maculatus and Guanapo haplotypes significantly.

### Trio binning and genome assembly

Trio binning (Koren et al. 2018) and haplotype assembly were performed on 170.1 Gb of PacBio long-read data using Canu (version 2.2, genomeSize=750 m). The contigs of both haplotypes were compared by dotplots (Supplemental Fig. S2) using D-Genies (Cabanettes and Klopp 2018).

BUSCO (version 4.12, reference "actinopterygii", revision odb10) and QUAST (Gurevich et al. 2013) were used for quality assessment of the assemblies.

Scaffolding of the assembled contigs to the published chromosome-level assembly (Fraser et al. 2020) was performed by RagTag (version 2.20-r1061) (Alonge et al. 2022). We used 11 Guanapo and nine Maculatus contigs >200 kb, respectively, for a primary assembly of $Y_{gua}$ (28.3 Mb) and $Y_{mac}$ (27.6 Mb) (see Supplemental Table S2; for details, see Supplemental Material). Telomere repeats are found at the start of the first contig Gua_tig00000738 and of the last contig Gua_tig000001309, but not on the corresponding Maculatus contigs or on the published X/Y Chr LR880656.1. Telomere repeats are also present on the unplaced XY scaffold_149F_0, which aligns to Gua_tig000001309 (Supplemental Fig. S3). Alignments for synteny assessment were visualized with AliTV (Ankenbrand et al. 2017).

An overview of the data flow and methods for genome assembly and data analysis is shown in Figure 6.

### Genome annotation

Protein coding genes on Y Chromosomes were annotated by synthesizing gene evidence from homologous alignments, transcriptome mapping, and ab initio prediction.

For homologous alignment, we collected 424,637 protein sequences from vertebrate database of Swiss-Prot (https://www.uniprot.org/statistics/Swiss-Prot), RefSeq database (proteins with ID starting with "NP" from "vertebrate_other"), and the NCBI genome annotation of human (GCF_000001405.39_GRCh38), zebrafish (GCF_000002035.6), platyfish (GCF_002775205.1), medaka (GCF_002234675.1), mummichog (GCF_011125445.2), turquoise killifish (GCF_001465895.1), and guppy (GCF_000633615.1). We aligned the protein sequences to the genome assembly using Exonerate (https://github.com/nathanweeks/exonerate) and GeneWise (Birney et al. 2004), respectively, to predict gene location and intron/exon structures. To speed up GeneWise, we used genBlastA to roughly locate the alignment first (She et al. 2009).

To collect transcriptome evidence, we downloaded the RNA-seq data from NCBI's BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA230881 (tissues include whole embryo, adult gonad, brain, tail, and pooled organs) and mapped them to the assembly using HISAT (Kim et al. 2015). StringTie was then used to interpret the mapping result for gene location and structure. In parallel processing, we assembled transcripts from the mapping result using Trinity (Haas et al. 2013) and aligned them to the genome using Splign (Kapustin et al. 2008).

We used AUGUSTUS for ab initio gene prediction (Stanke et al. 2006). AUGUSTUS was first trained by BUSCO with the

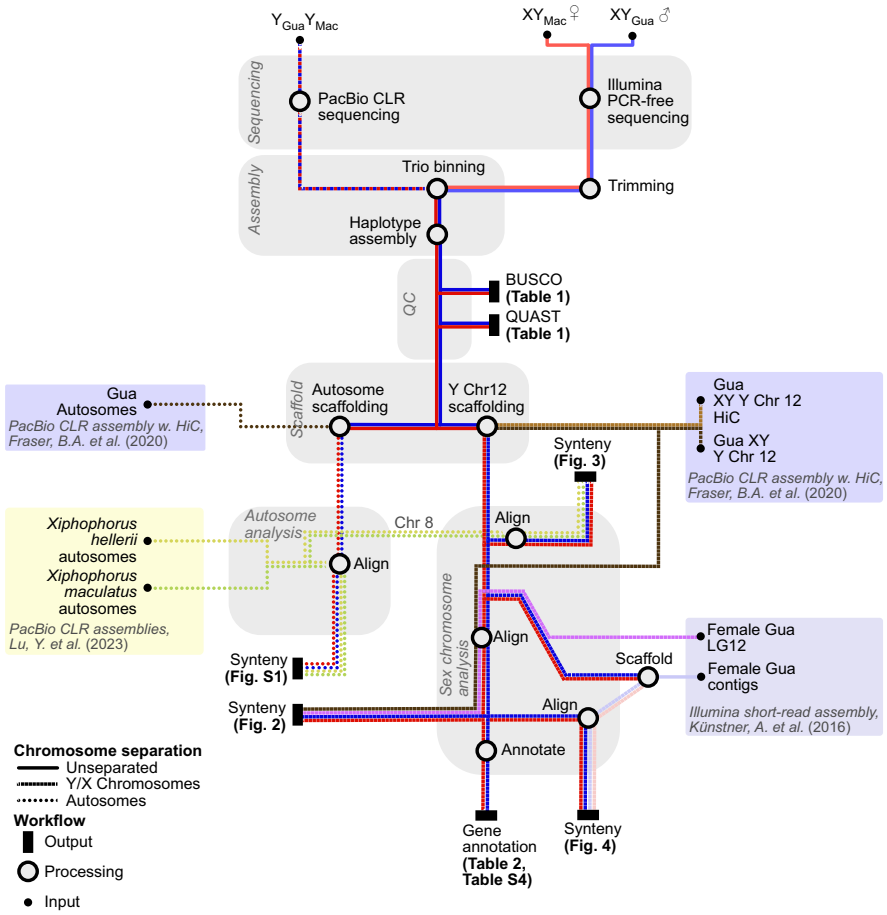


**Figure 6.** Overview of the data flow and methods for genome assembly and comparisons with published results.

parameter "-long" (Simão et al. 2015). We then trained it again using high-quality gene models that are commonly agreed upon by Exonerate, GeneWise, StringTie, and Splign. The retrained AUGUSTUS took the homologous and transcriptome gene models collected above as hints and screened the genome for ab initio gene prediction.

To synthesize all gene evidence into a final consistent set of annotation, we clustered overlapping homology gene models and kept the one best supported by transcriptome evidence (Supplemental Code 1), $f$. When the terminal exon of the kept gene model was poorly supported by transcriptome evidence, we screened the eliminated gene models for better supported terminal exons and replaced them. For genome regions where no homologous gene was predicted, ab initio gene models were recruited if they were fully supported by transcriptome evidence.

## Sequence alignment and divergence calculation

The sequences of the Y and X (NC_024342.1) Chromosomes were aligned using minimap2 (Li 2016, 2021), and the alignment was improved using Genome Alignment Tools from the Hiller laboratory as indicated below (Supplemental Code 2). Default parameters were used during minimap2 aligning. The alignment blocks were chained up using axtChain. The unaligned regions neighboring those blocks were realigned using patchChain.perl. We used RepeatFiller to incorporate the newly captured alignments into the alignment chain, whereas chainCleaner was used to remove obscure local alignments (Suarez et al. 2017; Osipova et al. 2019). At last, chainNet was used to collect alignment chains hierarchically to capture the orthologous alignments only (Kent et al. 2003). Based on the alignment, the sequence differences were then calculated as the percentage of SNPs and indels in 10 kb sliding windows along the alignment (Supplemental Code 3).

## RAD-tag sequencing and analysis of sex-specific markers

Of four different guppy strains kept in community tanks, 48 males and 48 females per strain were analyzed using the double digest RADSeq method (Poland et al. 2012), essentially as previously described (Kottler et al. 2015). Only 91 individuals of each strain could be successfully genotyped because five of the barcodes failed. The populations originated from the rivers Tranquille (West Trinidad), upper Quare (Quare_II 215-3), lower Oropouche (Oro209), East Trinidad, and Poza Azufre (PV6) Venezuela. These were all descendants of populations previously used for genotyping (Table 1 in Willing et al. 2010). The libraries were sequenced on an Illumina HiSeq 2000 instrument with 100 bp single-end reads. Reads were clustered with *Stacks* (Catchen et al. 2011), first by strain and then by sex. Alignment to male and female Guanapo WGA resulted in 3218 male-only tags (data set 1) (Fig. 5B, red bars).

In a second experiment, RAD-tag libraries were built from genomic DNA of 25 females and 25 males of the Caroni Swamp strain of *P. reticulata* (CS; WLC 3501) and sequenced on the HiSeq 2500 platform. The reads were demultiplexed and then determined present or not in each female and male individual in RADSex (R Core Team 2013; Feron et al. 2021). A tile plot describing the number of reads in the number of female/male individuals was then generated and used to reveal the sex-determination system of the species. Reads present only in males were then aligned to the genome to locate the sex-determination region (data set 2) (Fig. 5B, blue bars).

## PoolSex analysis

Male and female tissue samples pooled from 30 males and 30 females of *P. obscura* were sequenced with Illumina short reads, yielding the PoolSex reads for male and female. The reads were analyzed using the PoolSex pipeline (https://github.com/tankbuild/PoolSex). Specifically, male and female PoolSex reads were first mapped to the reference genome using BWA v0.7.17 with default parameters (Li 2014). Then the mapping coverage and content of sex-specific SNPs were determined in 50 kb sliding windows using PSASS (https://github.com/SexGenomicsToolkit/PSASS).

## TE identification

Transposable elements were identified using RepeatModeler2 (Flynn et al. 2020) and RepeatMasker (https://www.repeatmasker.org/). DNA sequences were first scanned by RepeatModeler for

de novo TE family identification. The results, together with models from FishTEDB (Shao et al. 2018; https://www.fishtedb.com), were transferred into RepeatMasker as a query library to further identify and mask TEs from the DNA sequences. The Kimura value of each TE was retrieved from the result file of RepeatMasker. TE density and Kimura distribution were plotted using ggplot2 in R (https://www.R-project.org/).

## Data access

The PacBio CLR sequencing data of the $Y_{gua}$ $Y_{mac}$ male, haplotype assemblies, and Pool-seq data have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession numbers PRJEB75520, PRJNA1111885, PRJNA1111886, and PRJNA1108343. The Illumina short-read data for PCR-free libraries of the Guanapo XY male and the Maculatus XY female have been submitted to the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home) under accession number PRJEB75519. The gene annotations on each haplotype are available at figshare (https://figshare.com/articles/dataset/Poecilia_reticulata_YY_chromosome_Guanapo_Maculatus_/21637295?file=38357789) and as Supplemental Data.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

This work was supported by a Gottfried Wilhelm Leibniz Award of the Deutsche Forschungsgemeinschaft and the Max Planck Society to D.W. K.D. and M.S. were supported by Texas State University, San Marcos, through the Xiphophorus Genetic Stock Center. We thank Indrajit Nanda for critical discussions, Pablo Carbonell for help and advice with trio binning, Joffrey Fitz for installing G-Browse and BLAST servers for this project, and Georg Schneider and Stefanja Topuz for fish care.

*Author contributions*: M.S., D.W., and C.D. designed the study and coordinated the work. D.W. and M.S. secured the funding. M.S. produced the YY male and provided DNA for RAD-tag and Pool-seq analyses. M.H. prepared the PacBio CLR library of the YY male and DNA libraries from Guanapo males and Maculatus females for Illumina sequencing. C.L. performed all sequencing. O.D. and I.B. did trio binning and DNA assembly. K.D. annotated the genomes and analyzed Y Chromosome structure. M.H. and Y.G. generated the RAD-tag and Pool-seq data. A.H. performed genomic PCR analysis. C.D. and M.S. wrote the manuscript with contributions from all other authors.

## References

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* **23:** 258. doi:10.1186/s13059-022-02823-7

Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. 2017. AliTV: interactive visualization of whole genome comparisons. *PeerJ Comput Sci* **3:** e116. doi:10.7717/peerj-cs.116

Bergero R, Charlesworth D. 2019. Reply to Wright et al.: how to explain the absence of extensive Y-specific regions in the guppy sex chromosomes. *Proc Natl Acad Sci* **116:** 12609–12610. doi:10.1073/pnas.1906633116

Bergero R, Gardner J, Bader B, Yong L, Charlesworth D. 2019. Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proc Natl Acad Sci* **116:** 6924–6931. doi:10.1073/pnas.1818486116

Birney E, Clamp M, Durbin R. 2004. Genewise and genomewise. *Genome Res* **14:** 988–995. doi:10.1101/gr.1865504

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120. doi:10.1093/bioinformatics/btu170

Braasch I, Schartl M, Volff J-N. 2007. Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BMC Evol Biol* **7:** 74. doi:10.1186/1471-2148-7-74

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

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. *Stacks*: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1:** 171–182. doi:10.1534/g3.111.000240

Chalopin D, Volff J-N, Galiana D, Anderson JL, Schartl M. 2015. Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Res* **23:** 545–560. doi:10.1007/s10577-015-9490-8

Charlesworth D, Bergero R, Graham C, Gardner J, Yong L. 2020. Locating the sex determining region of linkage group 12 of guppy (*Poecilia reticulata*). *G3 (Bethesda)* **10:** 3639–3649. doi:10.1534/g3.120.401573

Charlesworth D, Bergero R, Graham C, Gardner J, Keegan K. 2021a. How did the guppy Y chromosome evolve? *PLoS Genet* **17:** e1009704. doi:10.1371/journal.pgen.1009704

Charlesworth D, Graham C, Trivedi U, Gardner J, Bergero R. 2021b. PromethION sequencing and assembly of the genome of *Micropoecilia picta*, a fish with a highly degenerated Y chromosome. *Genome Biol Evol* **13:** evab171. doi:10.1093/gbe/evab171

Cimino MC. 1974. The nuclear DNA content of diploid and triploid *Poeciliopsis* and other poeciliid fishes with reference to the evolution of unisexual forms. *Chromosoma* **47:** 297–307. doi:10.1007/BF00328863

Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR, Larkin DM, Breden F, et al. 2019. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc Natl Acad Sci* **116:** 19031–19036. doi:10.1073/pnas.1905298116

Dor L, Shirak A, Kohn YY, Gur T, Weller JI, Zilberg D, Seroussi E, Ron M. 2019. Mapping of the sex determining region on linkage group 12 of guppy (*Poecilia reticulata*). *G3 (Bethesda)* **9:** 3867–3875. doi:10.1534/g3.119.400656

Du K, Pippel M, Kneitz S, Feron R, da Cruz I, Winkler S, Wilde B, Avila Luna EG, Myers G, Guiguen Y, et al. 2022. Genome biology of the darkedged splitfin, *Girardinichthys multiradiatus*, and the evolution of sex chromosomes and placentation. *Genome Res* **32:** 583–594. doi:10.1101/gr.275826.121

Du K, Ricci JMB, Lu Y, Garcia-Olazabal M, Walter RB, Warren WC, Dodge TO, Schumer M, Park H, Meyer A, et al. 2024. Phylogenomic analyses of all species of swordtail fishes (genus *Xiphophorus*) show that hybridization preceded speciation. *Nat Commun* **15:** 6609. doi:10.1038/s41467-024-50852-6

Feron R, Pan Q, Wen M, Imarazene B, Jouanno E, Anderson J, Herpin A, Journot L, Parrinello H, Klopp C, et al. 2021. RADSex: a computational workflow to study sex determination using restriction site-associated DNA sequencing data. *Mol Ecol Resour* **21:** 1715–1731. doi:10.1111/1755-0998.13360

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117:** 9451–9457. doi:10.1073/pnas.1921046117

Fong LJM, Darolti I, Metzger DCH, Morris J, Lin Y, Sandkam BA, Mank JE. 2023. Evolutionary history of the *Poecilia picta* sex chromosomes. *Genome Biol Evol* **15:** evad030. doi:10.1093/gbe/evad030

Fraser BA, Whiting JR, Paris JR, Weadick CJ, Parsons PJ, Charlesworth D, Bergero R, Bemm F, Hoffmann M, Kottler VA, et al. 2020. Improved reference genome uncovers novel sex-linked regions in the guppy (*Poecilia reticulata*). *Genome Biol Evol* **12:** 1789–1805. doi:10.1093/gbe/evaa187

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29:** 1072–1075. doi:10.1093/bioinformatics/btt086

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8:** 1494–1512. doi:10.1038/nprot.2013.084

Hallast P, Ebert P, Loftus M, Yilmaz F, Audano PA, Logsdon GA, Bonder MJ, Zhou W, Höps W, Kim K, et al. 2023. Assembly of 43 human Y chromosomes reveals extensive complexity and variation. *Nature* **621:** 355–364. doi:10.1038/s41586-023-06425-6

Kapustin Y, Souvorov A, Tatusova T, Lipman D. 2008. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* **3:** 20. doi:10.1186/1745-6150-3-20

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and

human genomes. *Proc Natl Acad Sci* **100:** 11484–11489. doi:10.1073/pnas.1932072100
Kersten S, Rabanal FA, Herrmann J, Hess M, Kronenberg ZN, Schmid K, Weigel D. 2023. Deep haplotype analyses of target-site resistance locus *ACCase* in blackgrass enabled by pool-based amplicon sequencing. *Plant Biotechnol J* **21:** 1240–1253. doi:10.1111/pbi.14033
Khoo G, Lim TM, Chan W-K, Phang VPE. 1999. Genetic basis of the variegated tail pattern in the guppy, *Poecilia reticulata*. *Zool Sci* **16:** 431–437. doi:10.2108/zsj.16.431
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12:** 357–360. doi:10.1038/nmeth.3317
Kirkpatrick M, Guerrero RF. 2014. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* **197:** 531–541. doi:10.1534/genetics.113.156026
Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Schartl M. 2022. Evolution of the canonical sex chromosomes of the guppy and its relatives. *G3 (Bethesda)* **12:** jkab435. doi:10.1093/g3journal/jkab435
Kirpichnikov VS. 1981. *Genetic bases of fish selection*. Springer-Verlag, Berlin.
Kitano J, Ansai S, Takehana Y, Yamamoto Y. 2024. Diversity and convergence of sex-determination mechanisms in teleost fish. *Annu Rev Anim Biosci* **12:** 233–259. doi:10.1146/annurev-animal-021122-113935
Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* **36:** 1174–1182. doi:10.1038/nbt.4277
Kottler VA, Künstner A, Koch I, Flötenmeyer M, Langenecker T, Hoffmann M, Sharma E, Weigel D, Dreyer C. 2015. *Adenylate cyclase 5* is required for melanophore and male pattern development in the guppy (*Poecilia reticulata*). *Pigment Cell Melanoma Res* **28:** 545–558. doi:10.1111/pcmr.12386
Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, Dreyer C. 2016. The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. *PLoS One* **11:** e0169087. doi:10.1371/journal.pone.0169087
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].
Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32:** 2103–2110. doi:10.1093/bioinformatics/btw152
Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37:** 4572–4574. doi:10.1093/bioinformatics/btab705
Lindholm A, Breden F. 2002. Sex chromosomes and sexual selection in poeciliid fishes. *Am Nat* **160 Suppl 6:** S214–S224. doi:10.1086/342898
Lisachov AP, Zadesenets KS, Rubtsov NB, Borodin PM. 2015. Sex chromosome synapsis and recombination in male guppies. *Zebrafish* **12:** 174–180. doi:10.1089/zeb.2014.1000
Long X, Charlesworth D, Qi J, Wu R, Chen M, Wang Z, Xu L, Fu H, Zhang X, Chen X, et al. 2023. Independent evolution of sex chromosomes and male pregnancy–related genes in two seahorse species. *Mol Biol Evol* **40:** msac279. doi:10.1093/molbev/msac279
Lu Y, Rice E, Du K, Kneitz S, Naville M, Dechaud C, Volff J-N, Boswell M, Boswell W, Hillier L, et al. 2023. High resolution genomes of multiple *Xiphophorus* species provide new insights into microevolution, hybrid incompatibility, and epistasis. *Genome Res* **33:** 557–571. doi:10.1101/gr.277434.122
McKinney GJ, Seeb JE, Pascal CE, Schindler DE, Gilk-Baumer SE, Seeb LW. 2020. Y-chromosome haplotypes are associated with variation in size and age at maturity in male chinook salmon. *Evol Appl* **13:** 2791–2806. doi:10.1111/eva.13084
Metzger DCH, Sandkam BA, Darolti I, Mank JE. 2021. Rapid evolution of complete dosage compensation in *Poecilia*. *Genome Biol Evol* **13:** evab155. doi:10.1093/gbe/evab155
Morris J, Darolti I, Bloch NI, Wright AE, Mank JE. 2018. Shared and species-specific patterns of nascent Y chromosome evolution in two guppy species. *Genes (Basel)* **9:** 238. doi:10.3390/genes9050238
Nanda I, Feichtinger W, Schmid M, Schröder JH, Zischler H, Epplen JT. 1990. Simple repetitive sequences are associated with differentiation of the sex chromosomes in the guppy fish. *J Mol Evol* **30:** 456–462. doi:10.1007/BF02101117
Nanda I, Schories S, Tripathi N, Dreyer C, Haaf T, Schmid M, Schartl M. 2014. Sex chromosome polymorphism in guppies. *Chromosoma* **123:** 373–383. doi:10.1007/s00412-014-0455-z
Nanda I, Schories S, Simeonov I, Adolfi MC, Du K, Steinlein C, Alsheimer M, Haaf T, Schartl M. 2022. Evolution of the degenerated Y-chromosome of the swamp guppy, *Micropoecilia picta*. *Cells* **11:** 1118. doi:10.3390/cells11071118
Osipova E, Hecker N, Hiller M. 2019. RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *GigaScience* **8:** giz132. doi:10.1093/gigascience/giz132
Paris JR, Whiting JR, Daniel MJ, Ferrer Obiol J, Parsons PJ, van der Zee MJ, Wheat CW, Hughes KA, Fraser BA. 2022. A large and diverse autosomal haplotype is associated with sex-linked colour polymorphism in the guppy. *Nat Commun* **13:** 1233. doi:10.1038/s41467-022-28895-4
Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* **7:** e32253. doi:10.1371/journal.pone.0032253
R Core Team. 2013. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
Schartl M, Larue L, Goda M, Bosenberg MW, Hashimoto H, Kelsh RN. 2016. What is a vertebrate pigment cell? *Pigment Cell Melanoma Res* **29:** 8–14. doi:10.1111/pcmr.12409
Shao F, Wang J, Xu H, Peng Z. 2018. FishTEDB: a collective database of transposable elements identified in the complete genomes of fish. *Database* **2018:** bax106. doi:10.1093/database/bax106
Sharma E, Künstner A, Fraser BA, Zipprich G, Kottler VA, Henz SR, Weigel D, Dreyer C. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genomics* **15:** 400. doi:10.1186/1471-2164-15-400
She R, Chu JS-C, Wang K, Pei J, Chen N. 2009. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res* **19:** 143–149. doi:10.1101/gr.082081.108
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210–3212. doi:10.1093/bioinformatics/btv351
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34:** W435–W439. doi:10.1093/nar/gkl200
Štundlová J, Hospodářská M, Lukšíková K, Voleníková A, Pavlica T, Altmanová M, Richter A, Reichard M, Dalíková M, Pelikánová Š, et al. 2022. Sex chromosome differentiation via changes in the Y chromosome repeat landscape in African annual killifishes *Nothobranchius furzeri* and *N. kadleci*. *Chromosome Res* **30:** 309–333. doi:10.1007/s10577-022-09707-3
Suarez HG, Langer BE, Ladde P, Hiller M. 2017. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33:** 1596–1603. doi:10.1093/bioinformatics/btx024
Tang L, Huang F, You W, Poetsch A, Nóbrega RH, Power DM, Zhu T, Liu K, Wang H-Y, Wang Q, et al. 2022. ceRNA crosstalk mediated by ncRNAs is a novel regulatory mechanism in fish sex determination and differentiation. *Genome Res* **32:** 1502–1515. doi:10.1101/gr.275962.121
Tripathi N, Hoffmann M, Weigel D, Dreyer C. 2009a. Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (*Poecilia reticulata*). *Genetics* **182:** 365–374. doi:10.1534/genetics.108.098541
Tripathi N, Hoffmann M, Willing E-M, Lanz C, Weigel D, Dreyer C. 2009b. Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proc Biol Sci* **276:** 2195–2208. doi:10.1098/rspb.2008.1930
van der Bijl W, Shu JJ, Goberdhan VS, Sherin LM, Cortázar-Chinarro M, Corral-López A, Mank JE. 2024. Deep learning reveals the role of copy number variation in the genetic architecture of a highly polymorphic sexual trait. bioRxiv doi:10.1101/2023.09.29.560175
Whiting JR, Paris JR, Parsons PJ, Matthews S, Reynoso Y, Hughes KA, Reznick D, Fraser BA. 2022. On the genetic architecture of rapidly adapting and convergent life history traits in guppies. *Heredity (Edinb)* **128:** 250–260. doi:10.1038/s41437-022-00512-6
Willing E-M, Bentzen P, van Oosterhout C, Hoffmann M, Cable J, Breden F, Weigel D, Dreyer C. 2010. Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Mol Ecol* **19:** 968–984. doi:10.1111/j.1365-294X.2010.04528.x
Winge Ö. 1922. One-sided masculine and sex-linked inheritance in *Lebistes reticulatus*. *J Genet* **12:** 145–162. doi:10.1007/BF02983078
Winge Ö. 1927. The location of eighteen genes in *Lebistes reticulata*. *J Genet* **18:** 1–43. doi:10.1007/BF03052599
Winge Ö. 1938. A lethal gene in the Y-chromosome of *Lebistes*. *Cr Trav Lab Carlsberg Ser Physiol* **22:** 203–210.
Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE. 2017. Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat Commun* **8:** 14251. doi:10.1038/ncomms14251
Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam BA, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE. 2019. On the power to detect rare recombination events. *Proc Natl Acad Sci* **116:** 12607–12608. doi:10.1073/pnas.1905555116

*Received May 16, 2024; accepted in revised form February 3, 2025.*