Research

Transposon wave remodeled the epigenomic landscape in the rapid evolution of X-Chromosome dosage compensation

David C.H. Metzger,[1] Imogen Porter,[1] Brendan Mobley,[2] Benjamin A. Sandkam,[3] Lydia J.M. Fong,[1] Andrew P. Anderson,[2] and Judith E. Mank[1]

[1]*Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada;* [2]*Biology Department, Reed College, Portland, Oregon 97202, USA;* [3]*Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA*

Sex chromosome dosage compensation is a model to understand the coordinated evolution of transcription; however, the advanced age of the sex chromosomes in model systems makes it difficult to study how the complex regulatory mechanisms underlying chromosome-wide dosage compensation can evolve. The sex chromosomes of *Poecilia picta* have undergone recent and rapid divergence, resulting in widespread gene loss on the male Y, coupled with complete X Chromosome dosage compensation, the first case reported in a fish. The recent de novo origin of dosage compensation presents a unique opportunity to understand the genetic and evolutionary basis of coordinated chromosomal gene regulation. By combining a new chromosome-level assembly of *P. picta* with whole-genome bisulfite sequencing and RNA-seq data, we determine that the YY1 transcription factor (YY1) DNA binding motif is associated with male-specific hypomethylated regions on the X, but not the autosomes. These YY1 motifs are the result of a recent and rapid repetitive element expansion on the *P. picta* X Chromosome, which is absent in closely related species that lack dosage compensation. Taken together, our results present compelling support that a disruptive wave of repetitive element insertions carrying YY1 motifs resulted in the remodeling of the X Chromosome epigenomic landscape and the rapid de novo origin of a dosage compensation system.

[Supplemental material is available for this article.]

Complete dosage compensation mechanisms have evolved in some organisms with highly diverged sex chromosomes to ameliorate the negative effects of haploinsuffiency (Ohno 1967; Bachtrog 2013). These complex mechanisms act across the entirety of the X Chromosome and represent a key model to study the evolution of coordinated transcriptional regulation (Kuroda et al. 2016) through the integration of sex-specific genomic and epigenomic processes (Ferrari et al. 2014). Understanding how dosage compensation mechanisms originate and evolve is important for understanding the molecular and evolutionary processes that regulate chromosome structure and gene regulatory processes (Dekker and Mirny 2016; Kuroda et al. 2016) and how the evolution of X-linked genes influences local adaptation (Lasne et al. 2017).

The extensive analysis of dosage compensation model systems has provided extraordinary discoveries on the genomic and epigenomic mechanisms involved in the regulation of gene expression at the whole-chromosome level, as different lineages have evolved different dosage compensation mechanisms. Although the specific mechanisms vary, a fundamental component of dosage compensation mechanisms studied to date is the up-regulation of genes on the X Chromosome to compensate for gene dosage imbalances in the heterogametic sex (Deng et al. 2011; Larsson et al. 2019; Lentini et al. 2022). In eutherian mammals, up-regulation of genes on the single active X Chromosome in both sexes is accompanied by inactivation of one X Chromosome in females. X inactivation is mediated by expression of the lncRNA *Xist* early during female development (Penny et al. 1996; Simon et al. 2013), ultimately leading to the transcriptional inactivation for most of the genes via the accumulation of epigenomic marks (Sahakyan et al. 2018). In *Drosophila*, dosage compensation of the X Chromosome in somatic cells is achieved through the male-specific lethal (MSL) complex, which acetylates histone H4, resulting in hyperexpression of X-linked genes in males (Lucchesi and Kuroda 2015). Substantial insight on how dosage compensation mechanisms are established on neosex chromosomes has been obtained from studying the spread of existing dosage compensation mechanisms to neosex chromosomes in *Drosophila* (Vicoso and Bachtrog 2009). However, it remains unclear how these complex adaptations to Y Chromosome degeneration can arise de novo, as the sex chromosomes of model systems are ancient, making it difficult to extrapolate the initial stages in the evolution of dosage compensation and obscuring cause from consequence.

A classic model of sex chromosome degeneration proposes loss of recombination results in the accumulation and acceleration of deleterious mutations on the nonrecombining chromosome. The evolution of complete dosage compensation is not a ubiquitous aspect of this theory, and there are relatively few examples in which sex chromosome divergence and gene loss require the acquisition of a complete dosage compensation mechanism

Corresponding author: dmetzger@zoology.ubc.ca
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.278127.123.

(Furman et al. 2020). Questions concerning how dosage compensation mechanisms evolve, whether dosage compensation evolves simultaneously with sex chromosome degradation, and whether dosage compensation facilitates gene loss remain unclear. One theory suggests that divergence of *cis*-regulatory regions in nonrecombining regions can cause rapid sex chromosome degeneration and acquisition of dosage compensation simultaneously (Lenormand et al. 2020). Because dosage compensation is thought to evolve during the earlier stages of sex chromosome divergence (Muyle et al. 2012; Martin et al. 2019), it has been difficult to test these theories, considering the age of dosage compensation mechanisms in model animal systems.

Recently, the first case of complete sex chromosome dosage compensation in fishes was identified in *Poecilia picta* (Darolti et al. 2019) and the close relative *Poecilia parae* (Metzger et al. 2021). *P. picta* and *P. parae* are closely related to the common guppy, *Poecilia reticulata*, and the Endler's guppy, *Poecilia wingei*. These species all share the same XY sex determination system in which Chromosome 12 is the sex chromosome, yet they show a remarkable diversity of X and Y Chromosome divergence and the evolution of complete dosage compensation (Fig. 1A). Although several phylogenetic models have been suggested for the origin of the sex chromosomes and dosage compensation in these species (Charlesworth et al. 2021a), comparative phylogenetic and genomic methods (Metzger et al. 2021; Darolti et al. 2023; Fong et al. 2023) suggest that the sex chromosomes diverged and complete dosage compensation evolved in a short interval between 14.8 and 18.5 mya (Fig. 1A; Metzger et al. 2021). Importantly, the sex chromosomes of *P. reticulata* and *P. wingei* lack any evidence of dosage compensation (Darolti et al. 2019), and karyogram data from *P. reticulata* and *P. wingei* show extensive homology with the X and Y relative to the highly degraded Y in *P. picta* (Nanda et al. 2022). Taken together, this system creates a powerful comparative genomic framework to study the early evolution of sex chromosomes. Additionally, as the only known case of complete X Chromosome dosage compensation in fish, it is unlikely that an existing dosage compensation system from a related species has been co-opted, presenting a unique opportunity to study the early de novo evolution of a novel dosage compensation system. In this study, we examine the genomic and epigenomic landscape of *P. picta* to identify sex-specific patterns consistent with X Chromosome dosage compensation, embedded within a comparative framework with related species.

Results

P. Picta genome assembly, annotation, and statistics

Our assembly of the *P. picta* genome is 744 Mb in total length, with benchmarking universal single-copy ortholog (BUSCO) completeness of 97.65% and an N50 of 33,053,486, representing a substantial improvement over previous genomic resources for the species (Charlesworth et al. 2021b). We recovered 23 chromosomes, numbered based on the syntenic comparison with the *P. reticulata* female genome assembly (Fig. 1B; Künstner et al. 2016), in which Chromosome 12 is the X Chromosome (Tripathi et al. 2009a,b). It is worth noting that this numbering system differs from some previous reports (Sandkam et al. 2021; Nanda et al. 2022). A total of 27,764 genes were annotated using RNA-seq data from head (with eyes removed), muscle, liver, testis, and ovary tissue, and 29.55% of the genome is classified as repeats (Supplemental Table 1).

Synteny analysis between our *P. picta* genome and the short-read Illumina-based female *P. reticulata* genome assembly (Künstner et al. 2016) identified several large apparent inversions and rearrangements between the assemblies on the autosomes and sex chromosome (Supplemental Fig. 1). In contrast, the *P. picta* genome is highly syntenic with the long-read male *P. reticulata* (Supplemental Fig. 2; Fraser et al. 2020) and *Xiphophorus helleri* (Supplemental Fig. 3; Schartl et al. 2013) genomes. These data suggest that the apparent inversions in the female *P. reticulata* genome could represent misassemblies from the short-read sequence data or polymorphic inversions in *P. reticulata*. Of particular relevance for this study are the several large inversions on the X Chromosome between *P. picta* and the male and female *P. reticulata* (Fig. 1C). Linkage analysis of the sex-determining region in *P. reticulata* has identified the sex-determining locus to the distal end of Chromosome 12 (Tripathi et al. 2009a); however, there is little agreement in the orientation of this region between the three assemblies. This disparity could be caused by a lack of sequence complexity and an abundance of repetitive and duplicated sequences in this region, making it difficult to sequence and assemble consistently. The *P. picta* X Chromosome (31.2 Mb) also has ~4 Mb at the distal end of the chromosome with very little homology with the X Chromosome in *P. reticulata* (26.61 Mb).

Complete dosage compensation is observed in male somatic tissues but not testis

Following the split from *P. reticulata* and *P. wingei* ~18.4 mya, the Y Chromosome in the common ancestor of *P. picta* and *P. parae* rapidly diverged from the X over a period of 3–4 Myr, leading to widespread gene loss and the rapid evolution of complete dosage compensation (Darolti et al. 2019; Metzger et al. 2021; Sandkam et al. 2021). In contrast, the Y Chromosome in *P. reticulata* and *P. wingei* has retained most of the coding content of the X, and therefore, there is little need for dosage compensation to evolve in these species (Wright et al. 2017; Darolti et al. 2019, 2020; Almeida et al. 2021).

The effectiveness of dosage compensation systems can vary between tissues (Nozawa et al. 2014). Previous studies of dosage compensation in *P. picta* and *P. parae* focused on the expression from samples consisting of mostly muscle tissue (Darolti et al. 2019; Metzger et al. 2021). To determine whether complete dosage compensation is equally effective across somatic and gonadal tissue or whether dosage compensation is restricted to specific tissues, we first analyzed RNA-seq data from male and female gonad, muscle, liver, and head tissue. Across all somatic tissue, we observe that the average gene expression level from the X is not statistically different between males and females and that the average gene expression level from the male X is not statistically different from male or female autosomes (Fig. 1D; for statistics, see Supplemental Table 2). However, the X expression was significantly lower in male gonad tissue compared with the average expression of male and female autosomal and female X genes in the gonad. This is consistent with a ubiquitous complete dosage compensation system acting in somatic tissues (Fig. 1D; Supplemental Table 2).

The capacity of dosage compensation mechanisms to buffer the effects of gene dosage imbalances can be limited for highly expressed genes (Vicoso and Bachtrog 2009; Harrison et al. 2012). To determine whether the efficiency of dosage compensation is affected by gene expression level, we analyzed dosage compensation patterns in four expression quantiles. Within each expression

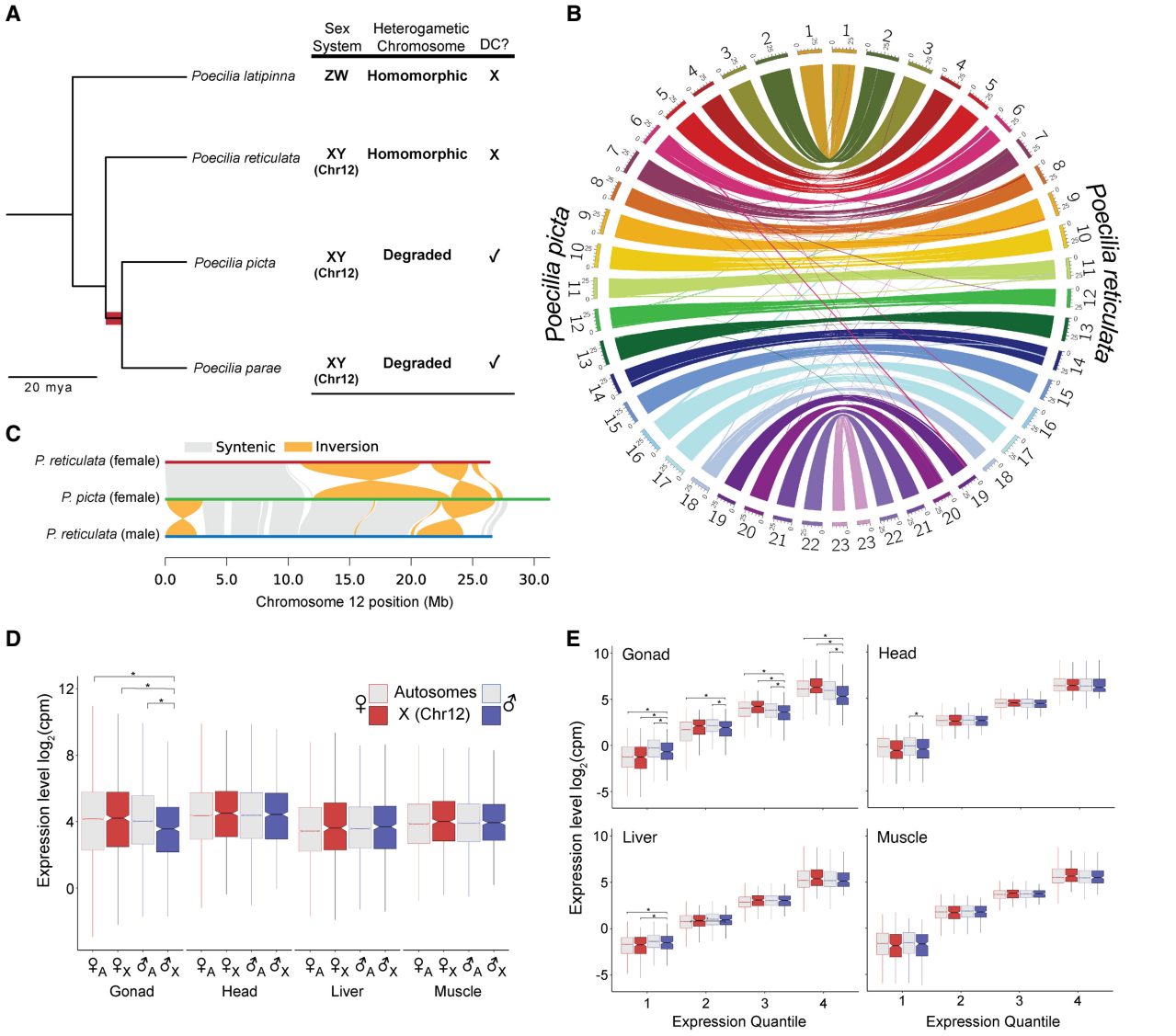

Figure 1. *Poecilia picta* genome assembly and analysis of tissue-specific dosage compensation. (*A*) Phylogeny of *Poecila* species depicting the diversity of XY heteromorphism between *P. reticulata*, *P. picta*, and *M. parae* and the ~3.7-Myr interval (shaded in red) for when the Y Chromosome degenerated and dosage compensation arose in the common ancestor of *P. picta* and *P. parae*. The phylogeny and divergence times are taken from The Fish Tree of Life (Rabosky et al. 2018) and adapted from Sandkam et al. (2021) and Metzger et al. (2021). (*B*) Sequence synteny between the female *P. picta* and the female *P. reticulata* genomes. Chromosomes are labeled 1–23 and are differentiated by color around the outer ring with *P. picta* chromosomes on the *left* and *P. reticulata* chromosomes on the *right*. The color of the connecting lines indicates the chromosomal origin of the sequence from the *P. picta* genome. The X Chromosome is Chromosome 12 and is highlighted in green. (*C*) Chromosome 12 alignments between the female *P. reticulata* (red horizontal line) (Künstner et al. 2016), our *P. picta* (green horizontal line), and the male *P. reticulata* (blue horizontal line) (Fraser et al. 2020) genome assemblies. Gray segments connecting horizontal lines represent syntenic regions. Orange segments connecting horizontal lines indicate inverted regions. (*D*) Comparison of autosomal (A; gray) and X Chromosome (X; red or blue fill) gene expression values in female (red) and male (blue) gonad, head, liver, and muscle tissues. Gene expression levels are presented as $\log_2$ counts per million (cpm) normalized to sequencing library size. The horizontal line of the box and whisker plot is the median; the box denotes the 25th and 75th percentile; and "whiskers" are 1.5 times the interquartile range. A significant difference ($P < 0.05$) in gene expression between groups is indicated by an asterisk, as determined by Bonferroni-corrected Wilcoxon rank-sum test. Complete list of *P*-values are available in Supplemental Table 2. (*E*) Comparison of gene expression levels between autosomal and X Chromosome genes in female and male gonad, head, liver, and muscle tissues in four expression quantiles. Colors and expression units are the same as in panel *B*. The complete list of *P*-values is available in Supplemental Table 3.

quantile, the average expression of sex chromosome genes in males is not significantly lower compared with male or female autosomal genes or female sex chromosome genes in somatic tissue, suggesting that dosage compensation is not affected by gene expression levels in *P. picta* (Fig. 1E; Supplemental Table 3). In contrast, X-linked genes were expressed at lower levels in the male gonad compared with autosomal genes and both sex chromosome and autosomal genes in females with the exception of the lowest expression quantile, in which male sex chromosome genes were more highly expressed. Significant differences in expression between X-linked genes in males and autosomal genes in the lowest expression quantile were also observed in the head and liver. These apparent differences in the expression of genes with the lowest expression may be an indication that dosage compensation is less

effective or does not act on genes with low expression (Deng et al. 2011).

DNA methylation in *P. picta*

Sex-specific epigenomic marks are commonly associated with facilitating sex-specific components of dosage compensation mechanisms, such as parent-of-origin-specific differences in DNA methylation (Golden et al. 2019) and maintenance of X inactivation in mammals (Riggs and Pfeifer 1992), or histone acetylation by the MSL complex in *Drosophila* (Conrad and Akhtar 2012). We therefore examined male and female muscle, liver, and gonad tissues for sex-specific DNA methylation patterns using WGBS data. Whole-genome CpG DNA methylation values are similar across all tissues (~72%), whereas methylation of other cytosine loci (CHG and CHH) is very low (≤2%) (Supplemental Table 4).

In vertebrates, CpG dinucleotides are not evenly distributed throughout the genome. CpG loci tend to be found in dense clusters in gene promoters called CpG islands (CGIs), and differential methylation of these CGIs plays an important role in gene regulation (Schmitz et al. 2019). Although this relationship between CGI methylation and gene expression regulation is thought to be conserved among vertebrates, the identification of CGIs and their role in gene expression in nonmodel systems is not well understood. In fishes, the abundance of CGIs is variable among distantly related species, and the characteristics that constitute CGIs (GC content, length and the CpG observed/expected ratio) differ between fishes and mammals, in which they are highly conserved among species (Han and Zhao 2008; Long et al. 2013). To gain a better understanding of the relationship between DNA methylation and genome function, we characterized the distribution of putative CGIs that we identified using a hidden Markov model relative to genic features. We found that 51% of putative gene promoters contained at least one CGI, and of the remaining promoters, 16% contained CGI shores (2 kb upstream of and downstream from a CGI) (Supplemental Table 5). These data are consistent with a role for promoter-associated DNA methylation in gene-specific regulation.

To further test this relationship, we divided our gene expression data into expression quantiles and calculated the average DNA methylation value for the putative promoter regions. We found that genes with the lowest expression had the highest promoter DNA methylation levels, whereas genes with higher expression had lower promoter methylation in somatic but not gonadal tissue (Supplemental Fig. 4). These data suggest the relationship between promoter methylation and gene expression is tissue-specific. It is possible that this pattern is the result of the breakdown of the relationship between CGIs and hypomethylated regions. To further explore this relationship, we characterized the distribution of CpG loci with high (>90%) and low (<10%) DNA methylation levels relative to CGI features. We found that a higher proportion of CpG loci with low methylation are located in CGIs compared with loci with high methylation across tissues, whereas CpG loci with high methylation were more abundant in repetitive elements (Supplemental Fig. 5). We also found that CGIs on the X Chromosome are more highly methylated compared with autosomal CGIs in both male and female liver tissues (Kruskal–Wallis $\chi^2 = 17.494$, df = 3, P-value = 0.0005592), and similar, marginally significant trends were observed in other tissues (muscle: Kruskal–Wallis $\chi^2 = 16.864$, df = 7, P-value = 0.01829; gonad: Kruskal–Wallis $\chi^2 = 9.4642$, df = 3, P-value = 0.02372) (Supplemental Fig. 6). Although these data suggest that hypomethylation of CGIs in gene promoters in *P. picta* is likely involved in tissue-specific gene regulation, the parameters of this relationship differ from those in mammals and require further exploration in other fish species.

Male-specific hypomethylation of the X Chromosome in *P. picta*

We next examined the WGBS data at the single-nucleotide level to identify site-specific patterns of sex-biased DNA methylation. Principal component analysis revealed that DNA methylation patterns in testis are distinct from other tissues (Supplemental Fig. 7). Male and female gonads show the greatest number of differentially methylated loci (DMLs; 744,852), and 79% of the DMLs are hypomethylated in the testis compared with the ovary (Supplemental Fig. 8; Supplemental Table 6). We found relatively fewer sex-biased DNA methylation sites between male and female somatic tissues (9585 DMLs in muscle; 7588 DMLs in liver). Consistent with gonad tissue, the majority of DMLs in somatic tissue are hypomethylated in male tissues (94% in muscle and 93% in liver) (Supplemental Table 6). Sex-specific DNA methylation patterns were significantly enriched on the X Chromosome in all tissues (Fishers's exact test, $P < 0.001$ in all tissue, gonad OR = 1.149912 [1.137536, 1.162376], muscle OR = 63.22567 [60.40327,66.14574], liver OR = 99.763 [93.6518, 106.2537]), particularly in somatic tissue (74% of DMLs in muscle and 82% of DMLs in the liver are located on the X compared with 5% in the gonad) (Supplemental Figs. 8, 9).

The methylation state of an individual CpG dinucleotide can be influenced by its proximity to other CpGs, in which sites in close proximity are likely to share the same methylation states. As a result, surrounding CpG sites tend to be coregulated and can form differentially methylated regions (DMRs) (Affinito et al. 2020). This phenomenon is particularly important in the context of allele-specific methylation, in which heterozygosity of a CpG disrupts the methylation potential at that site. Analysis of the DNA methylation state of individual CpGs can therefore be confounded by the presence of CpG single-nucleotide polymorphisms (SNPs). However, these CpG SNPs appear to have little effect when located in regions of differential methylation that are determined by *cis*-regulatory factors (Shoemaker et al. 2010).

Because the *P. picta* Y Chromosome is highly degraded and our reference genome was assembled from sequencing female tissue, it is possible that heterozygosity between the sex chromosomes is contributing to the identification of sex-biased methylation on the X Chromosome. To alleviate some of these effects, we next analyzed the WGBS data for sex-biased DMRs. We found that the genomic distribution of DMRs is consistent with the distribution of DMLs (Fig. 2; Supplemental Figs. 8, 9). We found 25,890 sex-biased DMRs between gonad tissue, 1264 DMRs in muscle tissue, and 330 DMRs in liver tissue. The majority of DMRs in somatic tissues are hypomethylated in males (79% in muscle and 83% in liver) (Supplemental Table 6) and are enriched on the sex chromosome (20% in muscle and 63% in liver; z-test P-value < 0.001), whereas the distribution of sex-specific DMRs in the gonad was equally hypo- and hypermethylated in gonad tissue (53% hypomethylated and 47% hypermethylated in males) and was evenly distributed among autosomes and the sex chromosome (z-test P-value = 0.0743) (Fig. 2; Supplemental Figs. 9, 10). Most DMRs were located in CGI shores and open sea regions and were located in intergenic regions (Supplemental Fig. 11).

To determine whether male-specific hypomethylated DMRs are involved in the localized regulation of individual genes, we analyzed the DNA methylation levels of CGIs in putative promoter regions. Despite finding an abundance of male-specific hypomethylated DMRs, we found no evidence for sex-biased DNA

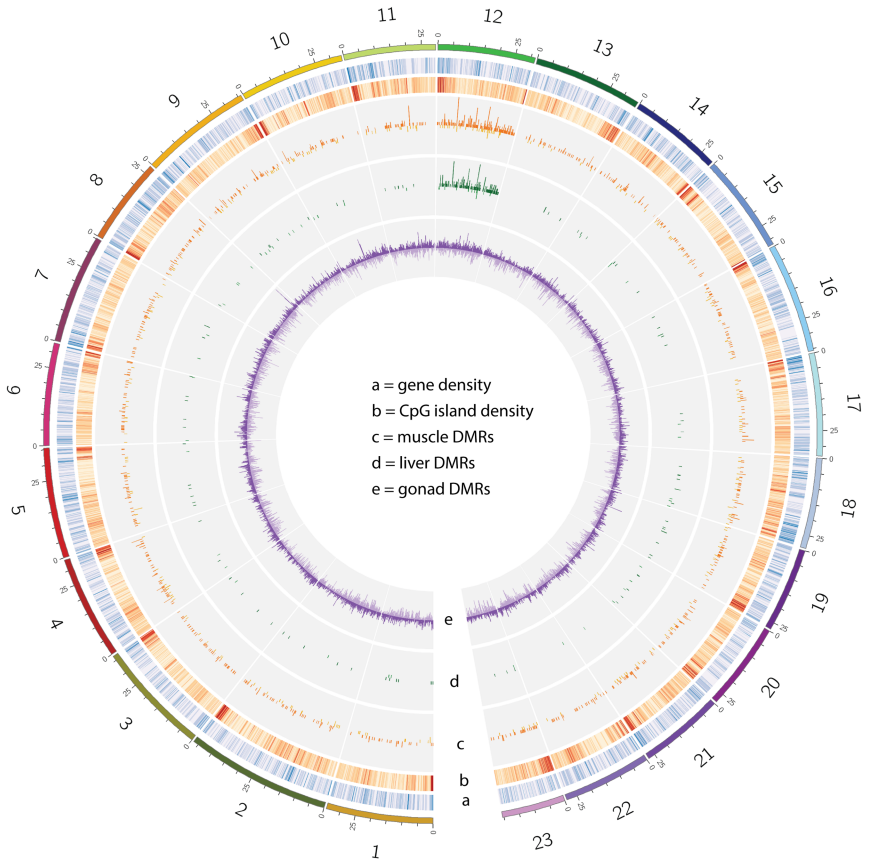

Figure 2. Distribution of sex-specific DMRs in *P. picta*. The *outer* ring of the Circos plot indicates chromosome number and megabase position. Chromosome colors are the same as in Figure 1. Tracks *a* and *b* are heatmaps depicting gene (blue) and CpG island (red) densities, each in 100-kb bins. Darker colors in the heatmap depict higher counts. Tracks *c*–*e* depict histograms of the number of differentially methylated regions (DMRs) in 100-kb bins in muscle (orange), liver (green), and gonad (purple) tissues. Histogram bars extending *outward* represent DMRs that are hypomethylated in males, and bars that extend toward the *center* of the plot represent DMRs that are hypomethylated in females. Chromosome 12 is the X Chromosome.

methylation of gene promoters (Supplemental Fig. 12), suggesting that the increased gene expression activity of male dosage compensation on the X Chromosome is not regulated by DNA methylation on a gene-by-gene basis.

If sex-biased DNA methylation patterns are involved in sex-biased up-regulation of gene expression and dosage compensation in males, we would not expect to see the same enrichment of X Chromosome hypomethylation when comparing patterns of tissue-specific methylation within a sex. In female tissues, we identified 51,656 DMRs between muscle and liver tissue, 56,645 DMRS between muscle and ovary tissue, and 49,001 DMRs between liver and ovary tissue, and these DMRs are distinct to each tissue (Supplemental Fig. 13). Tissue-specific DMRs in females were evenly distributed among autosomes and the X Chromosome (Supplemental Fig. 14). In male tissues, we identified 73,890 DMRs between muscle and liver tissue, 158,369 DMRS between muscle and testis tissue, and 84,805 DMRs between liver and testis tissue. Consistent with our observation of tissue-specific methylation in females, tissue-specific DMRs in males are distinct to each tissue (Supplemental Fig. 13). Although DMRs were evenly distributed among autosomes, there was a striking absence of tissue-specific DMRs on the X Chromosome in males, particularly between somatic tissues (Supplemental Fig. 15). These data suggest that DNA methylation of the male X is highly conserved among tissues and is distinct from female tissues.

Large-scale regulatory differences in the DNA methylation state of the X Chromosome between males and females could have karyotypic effects resulting in sex-specific differences in the heterochromatinization of the X, which typifies some examples of sex chromosome dosage compensation (Chadwick and Willard 2004; Rosin et al. 2022). We therefore examined male and female *P. picta* karyograms for evidence of chromosome-wide differences in heterochromatinization between male and female X Chromosomes. We identified, consistent with karyograms of *P. picta* from a Trinidadian population (Nanda et al. 2022), 46 acrocentric chromosomes, including one pair of heteromorphic XY sex chromosomes in males. We found the size of the X Chromosome to be very similar between males and females (Supplemental Fig. 16), suggesting that overall heterochromatinization of the X Chromosome does not differ substantially between the sexes.

Genomic regions with sex-biased DNA methylation are enriched for YY1 motifs

We next examined the DNA sequences of the DMRs for the possibility of enriched regulatory elements. Analysis of DMR sequences on the sex chromosome identified 35 putative regulatory elements enriched in regions with sex-biased DNA methylation. The 10 most significantly enriched motifs are presented in Table 1. The most significantly enriched regulatory element was the highly conserved DNA binding motif for the ubiquitously expressed YY1 transcription factor, which has a well-established role in sex-specific regulation of the *Xist* long noncoding RNA and dosage compensation in mammals (Do Kim et al. 2006; Donohoe et al. 2007; Jeon and Lee 2011; Makhlouf et al. 2014; Chen et al. 2016). The *P. picta* genome contains two YY1 orthologs, ANN22864 on Chr21 and ANN18583 on Chr22, that are ubiquitously expressed in head, muscle, liver, and gonad tissues (Supplemental Fig. 17). The DNA-binding domain of the *P. picta* YY1 protein sequence consists of four C2H2-type zinc fingers and is identical to the mammalian YY1 DNA-binding domain and the DNA-binding domain of other fishes (Supplemental Fig. 18). We also found a candidate binding motif for DPY-27, which is involved in dosage compensation in *C. elegans*. However, we were unable to identify an ortholog for DPY-27 in our annotation. The closest relative we could identify is SMC-4, which shares ~30% amino acid identity with DPY-27. Moreover, the motif identified in our analysis (wTGmGmwkGwAm) is substantially different from the conserved motif in *C. elegans* (nnCGCnnGGG) (Ercan et al. 2009), making it a far less compelling candidate.

Table 1. Motifs associated with DMRs on the *P. picta X Chromosome*

Motif	Target	*P*-value	DMRs with motif
	YY1	1×10^{-221}	53.07%
	DPY-27	1×10^{-174}	49.16%
	PB0162.1_Sfpi1_2	1×10^{-170}	43.02%
	MYB27	1×10^{-145}	37.99%
	ONECUT1	1×10^{-140}	36.87%
	PB0129.1_Glis2_2	1×10^{-139}	41.34%
	PB0028.1_Hbp1_1	1×10^{-137}	36.31%
	SPL9	1×10^{-137}	36.31%
	Ets1-distal(ETS)	1×10^{-132}	45.20%
	LIN-39	1×10^{-129}	41.90%

Table of the 10 most enriched transcription factor DNA binding motifs in differentially methylated regions between males and females. The motif column contains sequence logo plots of the DNA motifs identified by the enrichment analysis. The letter height represents the sequence conservation score in bits.

Further analysis of the YY1 DNA binding motif revealed that there are significantly more YY1 motifs on the *P. picta* sex chromosome compared with the autosomes (*P*-value < 0.001, *t*-value = 25.54, DF = 21), likely resulting from the replication of three of the six possible motifs (CCGCCATATTGT, CGGCCATTTTGT, and CAGCCA TTTTGT) on the sex chromosome relative to the autosomes (Fig. 3A). These motifs are highly conserved compared with mammalian YY1 motifs, and one in particular (CCGCCATATTGT) is identical to the YY1 binding motif of the *Xist* loci (CCGCCATnTT), which is regulated by sex-biased DNA methylation and is involved in mammalian dosage compensation (do Kim and Kim 2009). In total, we identified 2164 putative YY1 DNA binding motifs in the *P. picta* genome, and ~13% of all YY1 motifs identified are located in a DMR (liver = 281, muscle = 287, gonad = 288). Nearly all of the YY1 motifs located in a sex-biased DMR in somatic tissue were located on the X Chromosome (99%; liver = 279, muscle = 285), whereas 75% (226) of YY1 motifs in sex-biased gonad DMRs were on the X. Sex-biased DMRs containing YY1 motifs accounted for at least half of all sex-biased DMRs on the X (61% in liver, 63% in muscle, 50% in gonad) (Supplemental Table 7).

To determine whether the enrichment of YY1 motifs on the X Chromosome is present in other Poeciliids, we analyzed genomic sequences from several related species (Fig. 3A). We found, consistent with the pattern we observed in *P. picta*, significantly more YY1 sequence motifs on the X Chromosome of the sister species *P. parae* (*P*-value < 0.001, *t*-value = 4.228, DF = 21), which also shows complete X Chromosome dosage compensation (Metzger et al. 2021). Fold change differences in the enrichment of YY1 motifs on the X Chromosome between *P. picta* and *P. parae* are likely owing to differences in genome assembly quality. We found no enrichment of YY1 motifs on the X Chromosome of *P. reticulata* (*P*-value = 0.7166, *t*-value = −0.368, DF = 21), which shares the same X Chromosome but lacks substantial Y degeneration or dosage compensation (Darolti et al. 2019), or *P. latipinna* (*P*-value = 0.4929, *t*-value = −0.698, DF = 21), which has a ZW sex-determining mechanism located on a different chromosome from *P. picta* (Fig. 3A; Sola et al. 1992; Darolti et al. 2019).

Comparison of the YY1 motif distribution along the X Chromosome revealed a striking correlation between the distribution of DMRs and transposable elements (TEs) to YY1 motifs (*P* < 0.001, TE $R = 0.98$, muscle DMR $R = 0.59$, liver DMR $R = 0.61$, gonad DMR $R = 0.24$), particularly those in somatic (i.e., dosage compensated) tissues (Fig. 3B; Supplemental Fig. 19). The enrichment of YY1 motifs in these regions may also correlate with the accumulation of other DNA methylation–sensitive TF motifs. In this case, it would be more difficult to determine whether YY1 is the functional motif or whether the co-occurrence of YY1 with another motif is coincidental. To test this, we compared the abundance of YY1 motifs to the abundance of motifs enriched in DMRs. The two motifs with the highest correlation with YY1 were the DNA binding motifs for DPY-27 ($R^2 = 0.39$) and ROX1 ($R^2 = 0.35$) (Supplemental Fig. 20). These results suggest that the DNA binding motifs of DPY-27 and ROX1 may also have expanded on the X Chromosome along with the YY1 motif. As previously discussed, we were unable to identify an ortholog for DPY-27 in our annotation. Similarly, we were unable to identify an ortholog for ROX1 in our annotation. The lack of DPY-27 and ROX1 orthologs in the *P. picta* genome suggests that enrichment of these TF DNA binding motifs in DMRs is coincidental and that it is unlikely that these motifs recruit their designated TFs.

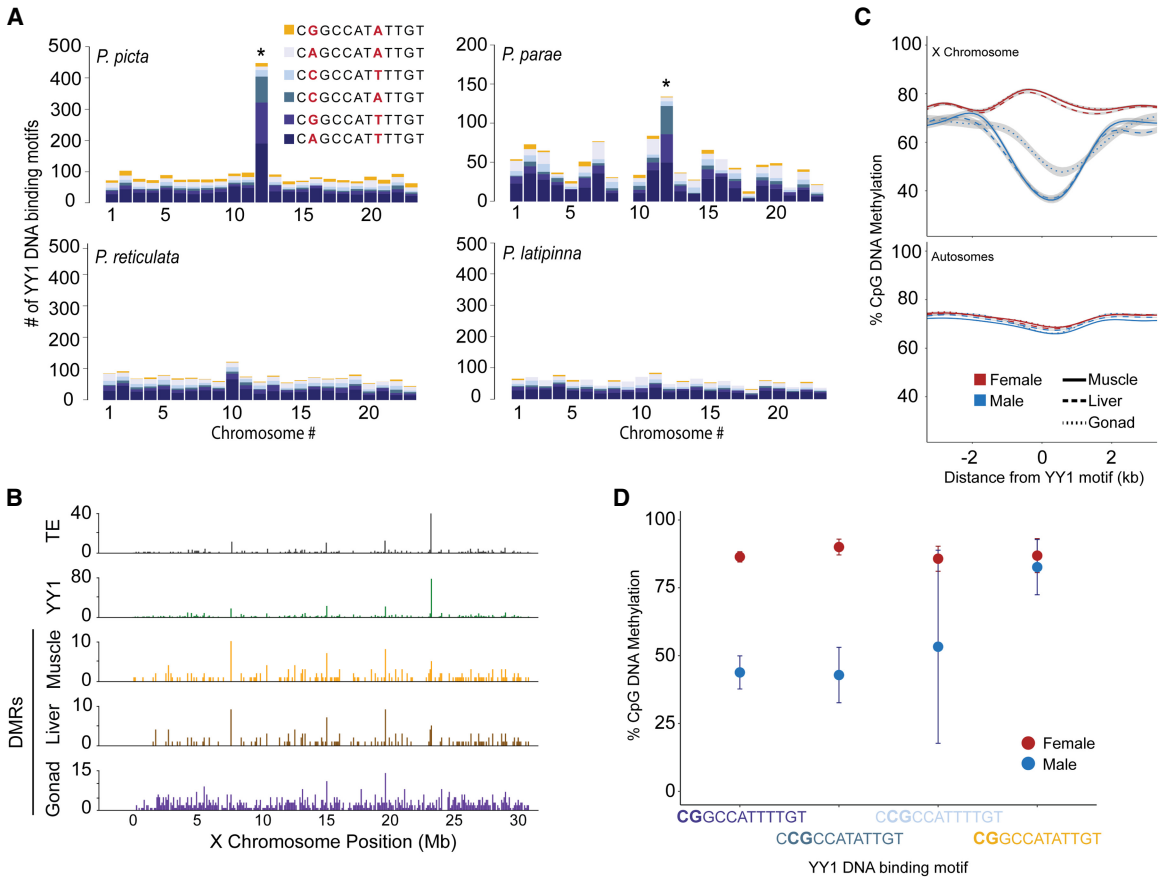

Figure 3. YY1 motif accumulated on the *P. picta* X Chromosome (Chromosome 12) and shows sex-specific DNA methylation. (*A*) Bar plots depict the chromosomal distribution of YY1 motifs. Colors represent the different YY1 motif sequences provided in Table 1. Nucleotides highlighted in red represent the nucleotide positions that vary among motifs. An asterisk denotes a significant enrichment in the total number of YY1 motifs on Chromosome 12 compared with autosomes based on results from a *t*-test with equal variance ($P < 0.05$). (*B*) Bar plots depicting the number of YY1 motif–containing transposable elements (TEs; gray), YY1 motif abundance (green), and male-specific hypomethylated regions in muscle (orange), liver (brown), and gonad (purple) tissue in 100-kb bins along the X Chromosome. (*C*) DNA methylation levels of genomic regions containing YY1 motifs ±2500 bp in females (red) and males (blue) on the X (*top*) and autosomes (*bottom*) presented as a loess regression (blue and red lines) with 95% confidence intervals in gray. Tissue-specific DNA methylation indicated by line type: Solid line indicates muscle; dashed line, liver; and dotted line, gonad. Zero marks the location of a YY1 motif. (*D*) Average DNA methylation values for CpG loci contained within YY1 motifs highlighted in large bold font on the *x*-axis. Data points represent the mean methylation value for the designated CpG loci for YY1 motifs on the sex chromosome in all tissues. Error bars represent 95% confidence intervals.

YY1 motifs have been shown to be regulated by differential DNA methylation patterns in other species (Do Kim et al. 2006; Cusack et al. 2020) and are known to play a role in *Xist*-mediated X Chromosome inactivation in eutherian mammals (Jeon and Lee 2011; Chapman et al. 2014; Makhlouf et al. 2014). To determine whether sex-biased DNA methylation patterns in *P. picta* could result in sex-biased YY1 binding to the regulatory motif, we analyzed DNA methylation levels around the YY1 motif on the sex chromosome and on the autosomes separately. We find that YY1 motifs on the X Chromosome are hypomethylated in all male tissues relative to females (Fig. 3C). The percentage difference in DNA methylation of YY1 motifs on the sex chromosome was similar in muscle (50%) and liver (48%) tissue. The DNA methylation of YY1 motifs in testis is hypomethylated compared with female tissues, but to a lesser extent (29%) compared with somatic tissues. In contrast, we observe no difference between male and female DNA methylation levels for YY1 motifs on the autosomes (Fig. 3C). Of the six putative YY1 binding sites, four contain CpG dinucleotides in the motif itself and could be directly influenced by DNA methylation. To test this, we compared DNA methylation values of the CpG loci for motifs located on the X Chromosome. Two of the four DNA binding motifs showed sex-biased methylation of the motif CpG, including the motif that is conserved with the *Xist* YY1 binding motif (CCGCCATATTGT) (Fig. 3D).

In *Drosophila* dosage compensation, the MSL complex acts locally and spreads out from high-affinity binding sites (HAS). Whole-chromosome dosage compensation is then achieved by the spatial proximity of HASs resulting from three-dimensional chromatin interactions (Ramírez et al. 2015). To determine whether a similar proximity effect occurs around YY1 motifs, we analyzed gene expression levels in the vicinity of YY1 motifs on the X Chromosome and on autosomes. We found no difference in the expression of genes around YY1 motifs between males and females in any tissue (Supplemental Fig. 21). These data suggest that dosage compensation efficiency is not affected by the proximity of genes to a YY1 binding. Moreover, the enrichment of male-biased hypomethylated regions on the *P. picta* sex chromosome is associated with genomic regions containing YY1 DNA binding motifs and not with regions containing gene promoters, suggesting that

these hypomethylated regions do not function on a gene-by-gene basis to upregulate genes on the male X. Taken together, these data suggest a putative dosage compensation system that involves the regulation of YY1 DNA-binding affinity via sex-specific DNA methylation. The DNA methylation–dependent binding affinity of the YY1 transcription factor could then result in sex-specific differences in the formation of long-range regulatory regions, such as altering 3D chromatin structure and the formation of insulated regions, resulting in global dosage compensation.

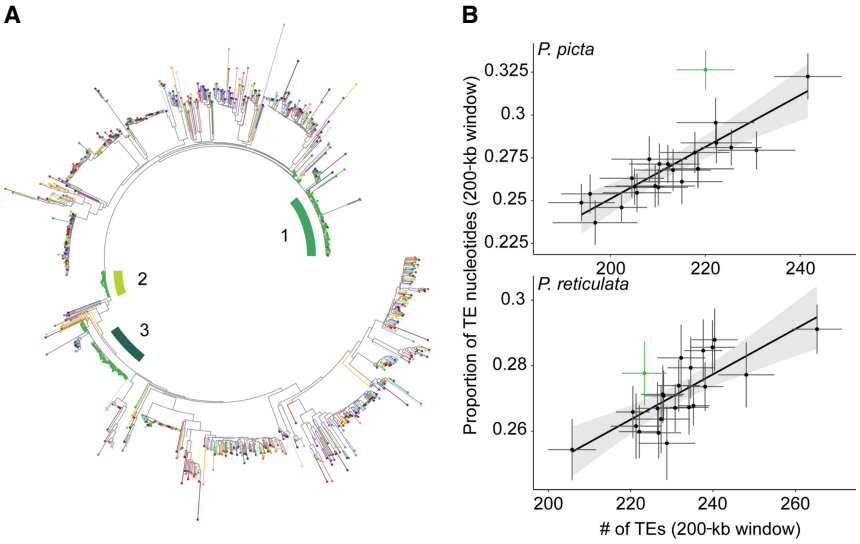

Figure 4. Recent expansion of repetitive elements containing YY1 motifs on the X Chromosome in *P. picta*. (*A*) Circular phylogeny of repetitive sequences containing at least one YY1 motif. Chromosomes are indicated by different colors as in Figures 1 and 2, with the X Chromosome in green. Green bars indicate the three clades where sequences are almost exclusively derived from the X Chromosome. (*B*) Relationship between the length and number of repetitive elements on each chromosome. Each point represents mean values in 200-kb bins for each chromosome (green indicates X). The proportion of TE nucleotides is calculated as the total number of TE nucleotides in a 200-kb bin. The number of TEs represents the total number of repetitive elements in 200-kb bins. Error bars are SE, and the shaded region is the 95% confidence interval line around the linear regression (solid black line).

Recent expansion of repetitive elements containing YY1 motifs

YY1 motifs can regulate the expression of TEs (Becker et al. 1993; Athanikar et al. 2004), providing a potential means for their replication and expansion throughout the genome. To test whether the enrichment of YY1 motifs on the *P. picta* X Chromosome can be attributed to TE replication, we isolated repetitive regions of the genome and identified those containing at least one YY1 motif for phylogenetic analysis. This phylogenetic analysis revealed that most of these unannotated repetitive regions from autosomal sequences do not cluster by chromosome. In contrast, sequences derived from the X Chromosome form three distinct clusters that are defined by very short internal branch lengths (Fig. 4A). The short internal branch lengths of these three clusters and their strict localization on the X strongly suggest that they originated from recent TE expansions. To further examine the repetitive element profile of the *P. picta* genome, we calculated Kimura distance values using 200-kb sliding windows for each chromosome. Of the 23 chromosomes, the sex chromosome contains an anomalous abundance of longer repeats (Fig. 4B) with low Kimura distance values (Supplemental Fig. 22), consistent with a recent expansion of repetitive elements on this chromosome, which may partially account for the size discrepancy of the X Chromosome assembly size between *P. picta* (31.3 Mb) and *P. reticulata* (26.6 Mb) (Fig. 1C).

TEs can be classified into two general categories: class I TEs or retrotransposons, which typically replicate via a "copy-and-paste" mechanism and require reverse transcription and an intermediate RNA phase, and class II TEs or DNA transposons, which typically replicate via a "cut-and-paste" mechanism facilitated by their encoded protein transposase. Given the distinct differences in replication pathways between class I and class II TEs, classification of the TE containing the YY1 motif can provide key insight into the mechanism of proliferation involved in the expansion of YY1 elements on the X Chromosome. Annotation of the consensus sequences from clades 1–4 depicted in Figure 4A suggests that the TE is a class II DNA transposon with some sequence similarity to a piggyBac-like element. Importantly, each consensus sequence contained at least one of the YY1 binding motifs that shows sex-biased DNA methylation (Supplemental Fig. 23). This result suggests that a "cut-and-paste" mechanism facilitated the recent expansion of TEs along the X Chromosome in *P. picta* and produced the rapid accumulation of the YY1 transcription factor motifs. However, the transposition of nonautonomous TEs like piggyBac elements is driven by a transposase encoded by its autonomous counterpart such that the piggyBac transposase would recognize and bind the cognate terminal inverted repeats (TIRs) and cause TTAA target site duplication (TSD) upon insertion. A nonautonomous transposon, if recently transposed, should display the same properties; however, because of the short length of the consensus sequences, we are unable to identify TIRs or the TTAA TSD, and thus, we are unable provide an unambiguous assignment to a specific superfamily.

Discussion

Understanding the structure, function, and origins of genomic processes is fundamental to understanding the mechanisms of genomic evolution. The diversity of sex chromosomes in Poeciliids coupled with the rapid evolution of a novel dosage compensation mechanism in *P. picta* provides a unique opportunity to understand the molecular mechanisms and evolutionary processes that result in the de novo evolution of a novel dosage compensation mechanism. From our high-quality, chromosome-level assembly of *P. picta*, we were able to identify key elements of the genomic and epigenomic architecture that led to the recent and rapid evolution of a novel dosage compensation system. Our results support the putative model that a recent wave of TE replication led to a sex chromosome–specific accumulation of YY1 DNA binding motifs and that sex-biased DNA methylation may promote male-specific activation of the YY1 motifs as a key component in establishing dosage compensation in *P. picta* (Fig. 5).

YY1 and dosage compensation in *P. picta*

YY1 is a ubiquitously expressed (Melé et al. 2015; Weintraub et al. 2017) DNA-binding zinc-finger transcription factor (Shi et al. 1991; Klenova et al. 1993) that has been shown to selectively bind to hypomethylated DNA sequences (Kim et al. 2003; Do Kim et al. 2006; Makhlouf et al. 2014; Yin et al. 2017; Cusack et al. 2020). In mammals, YY1 shares several characteristics with the conserved CCCTC-binding factor (CTCF) protein. Both CTCF and YY1 are ubiquitously expressed transcription factors that function as activators, repressors, and chromatin insulators and play important roles in imprinting (Donohoe et al. 2007). The CTCF protein has a well-established role in gene regulation and the formation of 3D chromatin organization such as topologically associated domains in mammals (Pérez-Rico et al. 2020). These gene expression domains are often demarcated by the formation of newly derived CTCF motifs resulting from a recent expansion of repetitive elements (Schmidt et al. 2012). Analogous to CTCF–CTCF interactions, YY1 dimerization has also been shown to promote the formation of DNA loops, forming insulated gene regulatory regions and facilitating enhancer–promoter interactions (López-Perrote et al. 2014; Saldaña-Meyer et al. 2014; Chen et al. 2016; Weintraub et al. 2017).

Our results are consistent with a recent expansion of YY1 motifs on the X Chromosome in the common ancestor of *P. picta* and *P. parae.* Overall, these data illustrate the convergence of three evolutionary processes. First, these data suggest a convergent putative model for the evolution of regulatory domains and a novel dosage compensation mechanism through a TE-mediated expansion of transcription factor motifs as seen in *Drosophila* (Matyunina et al. 2008; Ellison and Bachtrog 2013). Second, allele-specific expression patterns in *P. picta* and *P. parae* strongly suggest the convergent evolution of a "*Drosophila*-like" dosage compensation mechanism through a male-specific hyperexpression of the X Chromosome (Conrad and Akhtar 2012; Darolti et al. 2019; Metzger et al. 2021). Third, our study suggests convergent evolution with the mammalian dosage compensation system through co-opting the sex-specific DNA methylation of YY1 DNA-binding motifs. Sex-specific regulation of YY1 elements is well-established and plays a pivotal role in genomic imprinting and sex-specific regulation of the *Xist* long noncoding RNA involved X Chromosome silencing in mammals (Do Kim et al. 2006; Donohoe et al. 2007; Jeon and Lee 2011; Makhlouf et al. 2014; Chen et al. 2016). In the context of mammalian X inactivation, YY1 binding is regulated by the differences in the epigenetic landscape between the active and inactive X in females, where hypomethylation of the YY1 DNA-binding motif promotes YY1 binding and the expression of *Xist*.

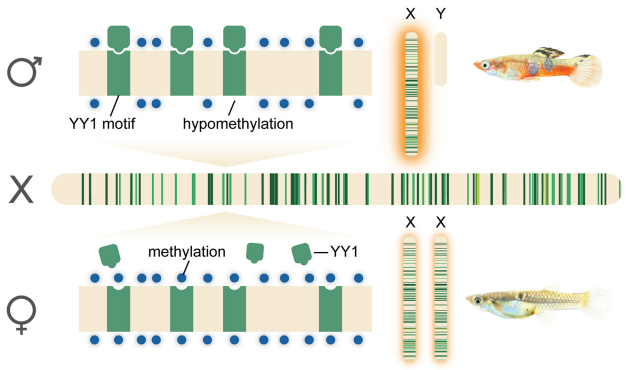

Figure 5. Proposed model for complete X Chromosome dosage compensation in *P. picta.* Hypomethylation of YY1 DNA binding motifs in males promotes male-specific binding of the YY1 transcription factor and the global up-regulation of genes on the X Chromosome in males. Hypermethylation of YY1 DNA motifs in females prevents YY1 from binding to the X, and thus, hyperexpression of X Chromosome genes is not observed in females. Green bars indicate position of YY1 motifs along the X Chromosome and correspond to the clade of repetitive elements that the YY1 motif is located in (based on colors in Fig. 4).

Although the exact role of YY1 in *P. picta* is not yet known, here we propose that male-specific hypomethylation of YY1 DNA binding motifs enhances YY1 binding affinity, resulting in male-specific hyperexpression of the X Chromosome (Fig. 5). Moreover, transcription of YY1 regulatory elements may produce a positive-feedback loop that could contribute to the stability of gene expression (Sigova et al. 2015) and maintain sex-specific activity of YY1 motifs once they are established. Future studies will provide key insights into the role of YY1 and the regulation of chromosome structure, enhancer–promoter interactions, and nucleation factors for the binding and spreading of dosage compensation machinery on the *P. picta* X Chromosome. Although some cases of sex chromosome dosage compensation are associated with heterochromatization (Chadwick and Willard 2004; Rosin et al. 2022), morphological analysis of metaphase spreads from males and females in *P. picta* showed no clear difference between male and female X Chromosome morphology. Comparison of male and female X Chromosomes using higher-resolution methods, such as C-banding, which can selectively stain heterochromatin-rich regions of individual chromosomes, may be more effective at identifying sex-specific differences in X Chromosome chromatin organization (Sumner 1972; Salvadori et al. 2015), or it may be that X Chromosome dosage compensation is achieved without differential heterochromatinization between males and females, as is the case in *Drosophila* (Zhou et al. 2013).

TE replication and sex chromosome evolution

TEs are major innovators of genome evolution and speciation. Replication and expansion of genomic regions via TE activity is an important process that contributes to the generation of genomic diversity such as the replication and formation of new TF binding sites (Schmidt et al. 2012; Xie et al. 2013; Sundaram and Wysocka 2020). In mammals, the accumulation of LINE-1 elements on the X Chromosome is thought to have played an important role in establishing X-inactivation, a process known as the Lyon "repeat hypothesis"(Lyon 1998; Bailey et al. 2000). In *Drosophila* neosex chromosomes (Marín et al. 1996), existing mechanisms of dosage compensation can spread with the expansion of binding motifs (Ellison and Bachtrog 2013; Zhou et al. 2013). However, many of the genes involved in the formation and spreading of the *Drosophila* MSL complex are not present in fish genomes. Annotation of the TE consensus sequences suggests that the TE containing the YY1 motif in *P. picta* is likely a class II DNA transposon with sequence similarity to a piggyBac-like element. The piggyBac superfamily of TEs belong to subclass I of DNA TEs, which replicate via a cut-and-paste mechanism that is prone to "local hopping" in the vicinity of the donor locus (Mátés et al. 2007). This localized replication provides a potential explanation for the enrichment of the TE containing the YY1 motif being restricted to the X Chromosome.

We have identified a recent repeat-associated expansion YY1-binding motifs on the *P. picta* X Chromosome, suggesting a potential model for the rapid evolution of dosage compensation. The X-

Chromosome enrichment of YY1 motifs and YY1-bearing TEs is absent from the syntenic chromosomes of *P. reticulata* and *Poecilia latipinna*, which both lack dosage compensation. This phylogenetic distribution suggests that the expansion occurred in the common ancestor to *P. picta* and *P. parae* ~18.4 mya in tandem with the rapid evolution of dosage compensation over 3–4 Myr (Fig. 1A; Metzger et al. 2021). Various models have been proposed for the evolutionary history of sex chromosomes in the clade of Poeciliids encompassing *P. reticulata*, *P. wingei*, *P. picta*, and *P. parae* (Darolti et al. 2019; Charlesworth et al. 2021a; Fong et al. 2023). However, all available evidence suggests that the sex chromosome arose in the ancestor to *P. reticulata*, *P. wingei*, *P. picta*, and *P. parae*, ~20 mya (Metzger et al. 2021; Sandkam et al. 2021; Darolti et al. 2023; Fong et al. 2023). Based on this model, extensive Y Chromosome degradation and dosage compensation evolved in the common ancestor of *P. picta* and *P. parae* after the split from *P. reticulata* and *P. wingei*.

Our data here are also consistent with the recent origin of the sex chromosomes, ~20 mya, and the rapid evolution of dosage compensation in the common ancestor of *P. picta* and *P. parae*. The enrichment of YY1 motifs is specific to the X Chromosome in *P. picta* and *P. parae*, the two species with complete sex chromosome dosage compensation. It is reasonable that the extensive sex chromosome divergence we observe in *P. picta* and *P. parae* was associated with the YY1-associated dosage compensation mechanism. These data suggest that the evolution of regulatory elements on the sex chromosome may facilitate the Y Chromosome degeneration and evolution of dosage compensation simultaneously (Lenormand et al. 2020). If the divergence and dosage compensation mechanism were much older and simply associated with sex chromosome turnover events in other species, as has been suggested by some (Charlesworth et al. 2021a), we would expect to observe the enrichment of YY1-motif bearing TEs on the syntenic regions in other related species, which we do not. Moreover, phylogenetic analysis of repetitive element sequences containing YY1 motifs supports the putative model of a recent TE wave that resulted in the accumulation of YY1 elements on the X, again consistent with a recent origin of dosage compensation.

Poecilia is the only known clade of fishes with complete X Chromosome dosage compensation. Aside from mammals, the only other vertebrate for which a complete dosage compensation mechanism has been proposed is in *Anolis* lizards (Marin et al. 2017). The extensive phylogenetic distance to other known mechanisms of whole-chromosome dosage compensation makes co-option of these mechanisms in *P. picta* highly unlikely. The dosage compensation system in *P. picta* therefore likely evolved largely de novo from a suite of genes and regulatory elements that acquired a novel role to selectively target and modulate expression of the X Chromosome in a sex-specific manner. Whether TE-mediated accumulation of YY1 motifs is the result of sex chromosome divergence or facilitated the rapid divergence between the X and Y requires further study. Here we present evidence for an elegant strategy to rapidly evolve a novel dosage compensation system through the recruitment of DNA binding motifs of a ubiquitously expressed transcription factor via waves of TE expansion coupled with sex-specific regulation of the YY1 regulatory motif via DNA methylation.

Methods

Fish collection and sampling for genome sequencing

Animals used in this study are from a laboratory-reared population of *P. picta* collected initially in the spring of 2019 from natural populations in Suriname and brought to the University of British Columbia aquatics facility, where they were kept in 20-L glass aquaria on a 12:12 day–night cycle at 26°C and 13 ppt salinity (Instant Ocean Sea Salt) and fed Hikari Fancy Guppy pellets supplemented with live brine shrimp daily. Before tissue sampling, individuals were euthanized using a lethal overdose of MS-222. For genome sequencing, muscular tail tissue was taken from the anal pore to the base of the pectoral fin from a single adult female. Tissue was immediately flash-frozen in liquid nitrogen and stored at −80°C. All work described here was performed under approval from the UBC Animal Care committee (permit A22-0239).

PacBio library and sequencing

Frozen tissue samples were shipped overnight on dry ice to the Dovetail Genomics production laboratory for Dovetail Omni-C library preparation and sequencing. DNA samples were quantified using a Qubit 2.0 fluorometer (Life Technologies). The Pacific Biosciences (PacBio) SMRTbell library (~20 kb) for PacBio sequel was constructed using a SMRTbell express template prep kit 2.0 (PacBio) using the manufacturer's recommended protocol. The library was bound to polymerase using the Sequel II binding kit 2.0 (PacBio) and loaded onto PacBio sequel II. Sequencing was performed on PacBio sequel II 8M SMRT cells.

De novo sequence assembly

The wtdbg2 long-read assembler (Ruan and Li 2020) was run with the following parameters: genome_size 0.7g --read_type sq --min_read_len 20000 --min_aln_len 8192. BlobTools v1.1.1 (Laetsch and Blaxter 2017) was used to identify potential contamination in the assembly based on BLAST (v2.9) results of the assembly against the NT database. A fraction of the scaffolds was identified as contaminant and was removed from the assembly. The filtered assembly was then used as an input to purge_dups v1.1.2 (Guan et al. 2020), and potential haplotypic duplications were removed from the assembly.

Dovetail Omni-C library preparation and sequencing

To generate the Omni-C libraries, chromatin was fixed with formaldehyde in the nucleus and then extracted. The extracted chromatin was then digested with DNase I. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. Crosslinks were then reversed and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeq X platform to produce approximately 40× sequence coverage. Reads with MQ > 50 were then used for scaffolding.

Scaffolding the assembly with HiRise

The input de novo assembly and Dovetail Omni-C library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail Omni-C library sequences were aligned to the draft input assembly using BWA (Li 2013). The separations of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed with HiRise to produce a likelihood model for genomic distance between read pairs. This model was used to join scaffolds, to identify and break putative misjoins, and to score

prospective joins. Genome completeness was estimated based on BUSCOs against the eukaryota_odb10 database (Simão et al. 2015).

Genome annotation

Coding sequences from *Poecilia formosa*, *P. latipinna*, *Poecilia mexicana*, *P. reticulata*, and *X. maculatus* were used to train the initial ab initio model for *P. picta* using the AUGUSTUS software (version 2.5.5). Six rounds of prediction optimization were performed with the software package provided by AUGUSTUS (Stanke et al. 2008). The same coding sequences were also used to train a separate ab initio model for *P. picta* using SNAP (version 2006-07-28). RNA-seq reads were mapped onto the genome using the STAR aligner software (version 2.7) (Dobin et al. 2013) and intron hints generated with the bam2hints tools within the AUGUSTUS software. MAKER (Cantarel et al. 2008), SNAP (Korf 2004), and AUGUSTUS (with intron–exon boundary hints provided from RNA-seq) were then used to predict for genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from *P. formosa*, *P. latipinna*, *P. mexicana*, *P. reticulata*, and *X. maculatus* to generate peptide evidence in the MAKER pipeline. Only genes that were predicted by both SNAP and AUGUSTUS were retained in the final gene sets. To help assess the quality of the gene prediction, AED scores were generated for each of the predicted genes as part of the MAKER pipeline. Genes were further characterized for their putative function by performing a BLAST search of the peptide sequences against the UniProt database. tRNA were predicted using the software tRNAscan-SE v2.05 (Chan and Lowe 2019).

Sequencing the mitochondrial genome

To sequence the mitochondrial genome, we extracted liver and brain from a male individual and enriched for mitochondria using a Qproteome mitochondria isolation kit (Qiagen). DNA was extracted using a QIAamp DNA micro kit (Qiagen) and sequenced on a MiSeq as 250-bp PE reads. Reads were trimmed using Trimmomatic (Bolger et al. 2014), resulting in 1.2 million sequences, which were de novo assembled using Geneious Prime (v.2022.2.1, Biomatters). The resulting mitochondrial sequence was aligned to sequences of close relatives (*P. parae*, *P. reticulata*, and *P. formosa*), and annotations were lifted over. A 245-bp insertion in the D-Loop of *P. picta* was confirmed using three Sanger reactions to sequence the 1.7-kb region spanning the insertion.

Synteny analysis

Sequence synteny was compared between the *P. picta* genome and the genomes of *P. reticulata* (obtained from the NCBI GenBank database [https://www.ncbi.nlm.nih.gov/genbank/] under accession numbers GCA_000633615.2 [female] and GCA_904066995.1 [male]) and *X. hellerii* (GenBank accession GCA_003331165.2). Syntenic regions were visualized using the output from satsuma2 in Circos (Krzywinski et al. 2009). Synteny between the *P. picta* and *P. reticulata* Chromosome 12 was visualized using plotsr (Goel and Schneeberger 2022). YY1 amino acid sequences were obtained from Ensembl (human = ENSP00000262238, mouse = ENSMUSP00000021692, pig = ENSSSCP00000053317, stickleback = ENSGACP00000011347, medaka = ENSORLP00000019577). Alignments between mammalian and fish YY1 amino acid sequences were performed in Geneious (v.2022.2.1, Biomatters).

Mitotic chromosome preparation

Cytogenetic analysis was conducted on adult fish (two male, two female) from the *P. picta* population described above. Mitotic chromosome preparations were obtained from gill arches according to the method previously described (Kligerman and Bloom 1977). Briefly, live specimens were exposed to a 0.01% colchicine solution for 7 h before being euthanized with a lethal overdose of MS-222 and dissected. Extracted organs were incubated in 0.4% KCl solution for 25–45 min and fixed with three changes of freshly prepared 3:1 ethanol:acetic acid fixative. Organs were minced in 50% acetic acid and mounted onto slides warmed to ~45°C and then stained for 10 min in a 5% Giemsa solution with phosphate buffer (pH 6.8) and air-dried. Slides were visualized on a Nikon Eclipse Ti-S inverted microscope (100×/1.30NA oil objective) equipped with a Veroptics IMX174 color camera (or cooled monochrome equivalent) with Micro-Manager software (MMStudio version 2.0.0, MMCore version 10.1.1). For each individual specimen, all visible chromosome spreads were photographed and evaluated for completeness. A minimum of eight high-quality images per specimen were selected for further analysis. Captured images were edited and analyzed with Adobe Photoshop (version 23.5.1).

Karyotype analysis

Chromosomes were sorted by length, grouped by morphology into homologous pairs, and classified according to the method of Levan et al. (2009). All chromosomes prepared from female specimens were successfully paired, whereas male spreads consistently contained two chromosomes that could not be paired in this manner. The two heteromorphic chromosomes in the male karyotype were designated as sex chromosomes based on morphological and size similarity of the larger unpaired chromosome to the female X. For each chromosome spread, a relative length index was calculated for the X Chromosome(s) by dividing the length of the X Chromosome(s) by the mean length of the chromosomes in the longest pair.

RNA-seq sampling and analysis

Adult liver, head (excluding the eyes), and whole-gonad tissue (somatic and gametic tissue) were sampled from adult male and virgin females from the sample laboratory population of *P. picta* described above immediately following a lethal overdose of MS-222. Tissue samples were immediately frozen in liquid nitrogen and stored at −80°C. A total of nine males and nine virgin females were sampled. For each tissue, we pooled three individuals of the same sex to create a total of three nonoverlapping male and three female sample pools for library preparation and sequencing. Frozen tissue samples were shipped on dry ice to Genewiz for RNA purification, mRNA library preparation, and sequencing.

Total RNA extraction was performed using the Qiagen RNeasy plus kit following the manufacturer's protocols. Total RNA was quantified using Qubit RNA assay and TapeStation 4200. Before library prep, RNA samples were DNase-treated followed by AMPure bead clean up and Qiagen FastSelect HMR rRNA depletion. Library preparation was performed with the NEBNext Ultra II RNA library prep kit following the manufacturer's protocols. Libraries were then sequenced on the NovaSeq 6000 platform in 2 × 150-bp configuration. RNA-seq data from muscle tissue (obtained from the NCBI Sequence Read Archive [SRA; https://www.ncbi.nlm.nih.gov/sra] under accession numbers SAMN29631966, SAMN29631977, SAMN29631978, SAMN29631989, SAMN31093659) (Fong et al. 2023) were also used for genome annotation and RNA-seq analysis.

Sequencing data were quality-filtered, and Illumina adapters were trimmed using Trimmomatic v0.36 (Bolger et al. 2014).

Reads were scanned with a four-base sliding window and discarded when the average Phred score was <15. Reads were trimmed to 120 bp, and the first 10 bp were trimmed from each read. Reads were mapped to the *P. picta* reference genome using HISAT2 v2.1 (Kim et al. 2019). Only uniquely mapped reads that aligned as pairs were retained for further analysis. The average mapping efficiency was 97.7% with an average mapped library size of 67,004,220 reads (Supplemental Table 8). Alignment BAM files were sorted and indexed using SAMtools v1.3.1. (Li et al. 2009) Aligned reads were assigned to genic regions using featureCounts v2.0.1 (Liao et al. 2014).

Gene expression analysis was conducted in R v4.1 (R core team 2022) using edgeR v3.34.1 (Robinson et al. 2010). Genes with zero counts in all samples were removed. We then filtered the data to retain only genes with more than 10 reads in each female or male sample. This filtered data set was then used to calculate normalization factors using the RLE method in the calcNormFactors() function. We estimated robust dispersions using the estimateDisp() followed by the estimateGLMRobustDisp () function. Normalized counts per million (cpm) values were then extracted from the data to compare gene expression levels between males and females and between genes located on the X Chromosome (Chromosome 12) compared with the autosomes. For the quantile expression analysis, the quantile() function from the R stats package was used to bin genes into the 0–25th, 25th–50th, 50th–75th, and 75th–100th percentiles based on gene expression values in males (Harrison et al. 2012). A Wilcoxon rank-sum test was used to identify differences in gene expression levels between groups using an FDR-corrected *P*-value <0.05. Gene promoters are defined as the region 2 kbp upstream of and 500 bp downstream from a transcription start site.

Bisulfite sequencing sampling and analysis

Liver, whole-gonad tissue (somatic and gametic tissue), and muscle tissue filets from the anal pore to the base of the tail (skin and bones removed) tissues were sampled from adult *P. picta* male and virgin females *P. picta* from the same laboratory population described above following a lethal overdose of MS-222. A total of 15 males and 15 females were sampled. Tissue was immediately flash-frozen in liquid nitrogen and stored at −80°C. To minimize the effect of genetic differences and developmental factors contributing to interindividual variation in DNA methylation patterns (Jaenisch and Bird 2003), we pooled tissues from five male or five female tissue samples to generate a total of three male and three female nonoverlapping sample pools for DNA isolation and sequencing with the exception of ovary tissue, for which we were only able to extract enough DNA from two of the three replicates. DNA was purified from pooled tissue samples using Qiagen DNeasy spin columns with on-column RNase A treatment following the manufacturer's protocol. Purified genomic DNA samples were frozen at −20°C and then shipped overnight on dry ice to the McGill University and Génome Québec Innovation Centre for bisulfite conversion, NEBNext enzymatic methyl-seq library preparation, and sequencing on the Illumina 6000 PE150 platform.

WGBS sequencing reads were quality-filtered, and adapters were trimmed using Trimmomatic v0.36 (Bolger et al. 2014). Reads were scanned with a four-base sliding window and discarded when the average Phred score was <15. Leading/trailing bases with a Phred score <3 were also removed. Trimmed and filtered reads were mapped to the *P. picta* reference genome using BSBolt (Farrell et al. 2021). Only uniquely mapped reads that mapped as pairs were retained for further analysis. The average mapping efficiency was 86.5%, and the average library size was 105,671,545 reads (Supplemental Table 9). BAM alignment files were prepared for duplicate removal using SAMtools v1.15.1 fixmate followed by SAMtools sort (Li et al. 2009). Duplicate reads were removed using SAMtools markdup and then indexed. We then used the BSBolt CallMethylation function to calculate DNA methylation levels at single-nucleotide resolution. Differential methylation analysis was conducted using the R package DSS v2.4 (Park and Wu 2016). DMLs and DMRs were identified for each tissue using the dmlTest() function with smoothing = TRUE followed by callDML() and callDMR(), respectively. CGIs were identified using a hidden Markov model using the makeCGI_1.3.4 R package (Wu et al. 2010). CGI shores are defined as the region 2 kb upstream of and downstream from a CGI; CGI shelves are 2–4 kb upstream of and downstream from a CGI; and open sea is the rest of the genome that is not contained within one of these three features. A Kruskal–Wallis test followed by a Dunn's post-hoc analysis was performed to test for differences between CGI methylation on the sex chromosome and autosomes.

Motif enrichment and YY1 analysis

Motif enrichment analysis for DMRs on the X Chromosome (Chromosome 12) was conducted using findMotifsGenome.pl in HOMER v4.11.1 (Heinz et al. 2010). Sequences from DMRs on the X Chromosome were used as the query sequences, and sequences from DMRs on the autosomes were used as background sequences. We used the matchPattern() function in the Biostrings v2.66.0 R package (https://bioconductor.org/packages/Biostrings) to search for YY1 motifs in the *P. picta* genome and to also identify repetitive regions containing YY1 motifs. Enrichment of the other DNA motifs was performed using the consensus motifs from the HOMER output files (Supplemental Table 10) using the SeqKit locate function (Shen et al. 2016). Significant enrichment of DNA-binding motifs on the X Chromosome was determined using a paired *t*-test with equal variance ($P<0.05$) and using motif count data normalized to chromosome length. Because the *P. latipinna* genome consists of unplaced scaffolds, the chromosome length values for *P. reticulata* were used to normalize the YY1 motive count values for *P. latipinna*. The DNA methylation of YY1 motifs was calculated as the mean DNA methylation value of CpG loci ± 500 bp from a YY1 motif.

De novo repeat element analysis

For the de novo identification of repetitive elements in the *P. picta* genome we used the RepeatModeler v2.0.1 (Flynn et al. 2020) to generate a repeat element library. We then used RepeatMasker V4.1.2 (Smit et al. 2013–2015) to annotate the repeat elements in the *P. picta* genome and to calculate kimura divergence estimates. Repetitive elements containing YY1 motifs were aligned using MUSCLE v.5.1 (Edgar 2004). The output file from the MUSCLE alignment was then processed using FastTree V2.1 (Price et al. 2009). The phylogeny from FastTree was visualized using the ggtree v3.0.4 package (Yu et al. 2017) in R. To compare TE length and abundance between chromosomes, we calculated the total length (in base pairs) and number of TEs in 200-kb nonoverlapping sliding windows. The mean total length and count per 200 kb for each scaffold were then plotted against each other in a scatter plot with standard error bars and a linear regression line. Annotation of the TE in clades 1, 2, and 3 from the phylogenetic analysis (Fig. 4A) was performed by aligning the sequences from the three clades separately using MUSCLE v.5.1 to obtain three consensus sequences (one for each clade). The consensus sequences were then annotated using Classify TE (Panta et al. 2021).

Data access

The *P. picta* genome generated in this study has been submitted to the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) under accession number JAVYJU000000000. The sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accessions PRJNA862953, PRJNA884377, and PRJNA884372. Source code generated in this study is available as Supplemental Code and has also been submitted to GitHub (https://github.com/manklab/Metzger_et_al_picta_DC). The genome assembly and annotation files can be downloaded from the Dryad repository (https://doi.org/10.5061/dryad.5qfttdzcf).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Jacelyn Shu for the scientific illustrations in Figure 4, for help with animal husbandry, and for tissue dissections for karyotype analysis. We thank Clara Lacy for the drawings of male and female *P. picta*. We also thank Wouter van der Bijl for assistance with R and members of the Mank laboratory and three anonymous reviewers for helpful and constructive suggestions on this manuscript. This work was funded by grants from the European Research Council (grant no. 680951), Natural Sciences and Engineering Research Council of Canada (NSERC), and Canada Foundation for Innovation, as well as a Canada 150 Research Chair to J.E.M. We also acknowledge funding from a Dovetail Genomics Tree of Life Award to D.C.H.M. and funding from NSERC Undergraduate Student Research Award grant awarded to I.P.

Author contributions: D.C.H.M. and J.E.M. conceived the study and wrote the manuscript. D.C.H.M. performed tissue collection and sample preparation for genome sequencing, WGBS, and RNA-seq data; performed bioinformatic analysis of WGBS and RNA-seq data; motif enrichment; and characterization of YY1 repetitive elements. I.P. identified and characterized repetitive elements in the *P. picta* genome. B.M. and A.P.A. performed the karyotype analysis. B.A.S. sequenced the mitochondrial genome, performed YY1 amino acid alignments, and helped with sample collection. L.J.M.F. helped with sample collection.

References

Affinito O, Palumbo D, Fierro A, Cuomo M, De Riso G, Monticelli A, Miele G, Chiariotti L, Cocozza S. 2020. Nucleotide distance influences co-methylation between nearby CpG sites. *Genomics* **112:** 144–150. doi:10.1016/j.ygeno.2019.05.007

Almeida P, Sandkam BA, Morris J, Darolti I, Breden F, Mank JE. 2021. Divergence and remarkable diversity of the Y chromosome in guppies. *Mol Biol Evol* **38:** 619–633. doi:10.1093/molbev/msaa257

Athanikar JN, Badge RM, Moran JV. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* **32:** 3846–3855. doi:10.1093/nar/gkh698

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14:** 113–124. doi:10.1038/nrg3366

Bailey JA, Carrel L, Chakravarti A, Eichler EE. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci* **97:** 6634–6639. doi:10.1073/pnas.97.12.6634

Becker KG, Swergold G, Ozato K, Thayer RE. 1993. Binding of the ubiquitous nuclear transcription factor YY1 to a *cis* regulatory sequence in the human LINE-1 transposable element. *Hum Mol Genet* **2:** 1697–1702. doi:10.1093/hmg/2.10.1697

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

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18:** 188–196. doi:10.1101/gr.6743907

Chadwick BP, Willard HF. 2004. Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci* **101:** 17450–17455. doi:10.1073/pnas.0408021101

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* **1962:** 1–14. doi:10.1007/978-1-4939-9173-0_1

Chapman AG, Cotton AM, Kelsey AD, Brown CJ. 2014. Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. *BMC Genet* **15:** 89. doi:10.1186/s12863-014-0089-4

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

Chen CY, Shi W, Balaton BP, Matthews AM, Li Y, Arenillas DJ, Mathelier A, Itoh M, Kawaji H, Lassmann T, et al. 2016. YY1 binding association with sex-biased transcription revealed through X-linked transcript levels and allelic binding analyses. *Sci Rep* **6:** 37324. doi:10.1038/srep37324

Conrad T, Akhtar A. 2012. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13:** 123–134. doi:10.1038/nrg3124

Cusack M, King HW, Spingardi P, Kessler BM, Klose RJ, Kriaucionis S. 2020. Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. *Genome Res* **30:** 1393–1406. doi:10.1101/gr.257576.119

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

Darolti I, Wright AE, Mank JE. 2020. Guppy Y chromosome integrity maintained by incomplete recombination suppression. *Genome Biol Evol* **12:** 965–977. doi:10.1093/gbe/evaa099

Darolti I, Fong LJM, Sandkam BA, Metzger DCH, Mank JE. 2023. Sex chromosome heteromorphism and the fast-X effect in poeciliids. *Mol Ecol* **32:** 4599–4609. doi:10.1111/mec.17048

Dekker J, Mirny L. 2016. The 3D genome as moderator of chromosomal communication. *Cell* **164:** 1110–1121. doi:10.1016/j.cell.2016.02.007

Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR, et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet* **43:** 1179–1185. doi:10.1038/ng.948

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

do Kim J, Kim J. 2009. YY1's longer DNA-binding motifs. *Genomics* **93:** 152–158. doi:10.1016/j.ygeno.2008.09.013

Do Kim J, Hinz AK, Bergmann A, Huang JM, Ovcharenko I, Stubbs L, Kim J. 2006. Identification of clustered YY1 binding sites in imprinting control regions. *Genome Res* **16:** 901–911. doi:10.1101/gr.5091406

Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. 2007. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* **25:** 43–56. doi:10.1016/j.molcel.2006.11.017

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5:** 113. doi:10.1186/1471-2105-5-113

Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342:** 846–850. doi:10.1126/science.1239552

Ercan S, Dick LL, Lieb JD. 2009. The *C. elegans* dosage compensation complex propagates dynamically and independently of X chromosome sequence. *Curr Biol* **19:** 1777–1787. doi:10.1016/j.cub.2009.09.047

Farrell C, Thompson M, Tosevska A, Oyetunde A, Pellegrini M. 2021. Bisulfite bolt: a bisulfite sequencing analysis platform. *GigaScience* **10:** giab033. doi:10.1093/gigascience/giab033

Ferrari F, Alekseyenko AA, Park PJ, Kuroda MI. 2014. Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nat Struct Mol Biol* **21:** 118–125. doi:10.1038/nsmb.2763

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:** 5. 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

Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. 2020. Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* **12:** 750–763. doi:10.1093/gbe/evaa081

Goel M, Schneeberger K. 2022. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38:** 2922–2926. doi:10.1093/bioinformatics/btac196

Golden LC, Itoh Y, Itoh N, Iyengar S, Coit P, Salama Y, Arnold AP, Sawalha AH, Voskuhl RR. 2019. Parent-of-origin differences in DNA methylation of X chromosome genes in T lymphocytes. *Proc Natl Acad Sci* **116:** 26779–26787. doi:10.1073/pnas.1910072116

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36:** 2896–2898. doi:10.1093/bioinformatics/btaa025

Han L, Zhao Z. 2008. Comparative analysis of CpG islands in four fish genomes. *Comp Funct Genomics* **2008:** 565631. doi:10.1155/2008/565631

Harrison PW, Mank JE, Wedell N. 2012. Incomplete sex chromosome dosage compensation in the Indian meal moth, *Plodia interpunctella*, based on de novo transcriptome assembly. *Genome Biol Evol* **4:** 1118–1126. doi:10.1093/gbe/evs086

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33:** 245–254. doi:10.1038/ng1089

Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146:** 119–133. doi:10.1016/j.cell.2011.06.026

Kim J, Kollhoff A, Bergmann A, Stubbs L. 2003. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, *Peg3*. *Hum Mol Genet* **12:** 233–245. doi:10.1093/hmg/ddg028

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37:** 907–915. doi:10.1038/s41587-019-0201-4

Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanenkov VV. 1993. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-*myc* gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol* **13:** 7612–7624. doi:10.1128/mcb.13.12.7612-7624.1993

Kligerman AD, Bloom SE. 1977. Rapid chromosome preparations from solid tissues of fishes. *J Fish Res Board Canada* **34:** 266–269. doi:10.1139/f77-039

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* **5:** 59. doi:10.1186/1471-2105-5-59

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645. doi:10.1101/gr.092759.109

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

Kuroda MI, Hilfiker A, Lucchesi JC. 2016. Dosage compensation in *Drosophila*: a model for the coordinate regulation of transcription. *Genetics* **204:** 435–450. doi:10.1534/genetics.115.185108

Laetsch DR, Blaxter ML. 2017. Blobtools: interrogation of genome assemblies. *F1000Res* **6:** 1287. doi:10.12688/f1000research.12232.1

Larsson AJM, Coucoravas C, Sandberg R, Reinius B. 2019. X-chromosome upregulation is driven by increased burst frequency. *Nat Struct Mol Biol* **26:** 963–969. doi:10.1038/s41594-019-0306-y

Lasne C, Sgrò CM, Connallon T. 2017. The relative contributions of the X Chromosome and autosomes to local adaptation. *Genetics* **205:** 1285–1304. doi:10.1534/genetics.116.194670

Lenormand T, Fyon F, Sun E, Roze D. 2020. Sex chromosome degeneration by regulatory evolution. *Curr Biol* **30:** 3001–3006.e5. doi:10.1016/j.cub.2020.05.052

Lentini A, Cheng H, Noble JC, Papanicolaou N, Coucoravas C, Andrews N, Deng Q, Enge M, Reinius B. 2022. Elastic dosage compensation by X-chromosome upregulation. *Nat Commun* **13:** 1854. doi:10.1038/s41467-022-29414-1

Levan A, Fredga K, Sandberg AA. 2009. Nomenclature for centromeric chromosome position on chromosomes. *Hereditas* **52:** 201–220. doi:10.1111/j.1601-5223.1964.tb01953.x

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN]. http://arxiv.org/abs/1303.3997

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:** 2078–2079. doi:10.1093/bioinformatics/btp352

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grützner F, Odom DT, Patient R, Ponting CP, et al. 2013. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife* **2:** e00348. doi:10.7554/eLife.00348

López-Perrote A, Alatwi HE, Torreira E, Ismail A, Ayora S, Downs JA, Llorca O. 2014. Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J Biol Chem* **289:** 22614–22629. doi:10.1074/jbc.M114.567040

Lucchesi JC, Kuroda MI. 2015. Dosage compensation in *Drosophila*. *Cold Spring Harb Perspect Biol* **7:** a019398. doi:10.1101/cshperspect.a019398

Lyon MF. 1998. X-Chromosome inactivation: a repeat hypothesis. *Cytogenet Genome Res* **80:** 133–137. doi:10.1159/000014969

Makhlouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C. 2014. A prominent and conserved role for YY1 in *Xist* transcriptional activation. *Nat Commun* **5:** 4878. doi:10.1038/ncomms5878

Marín I, Franke A, Bashaw GJ, Baker BS. 1996. The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* **383:** 160–163. doi:10.1038/383160a0

Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, Liechti A, Halbert J, Brüning T, Mössinger K, et al. 2017. Convergent origination of a *Drosophila*-like dosage compensation mechanism in a reptile lineage. *Genome Res* **27:** 1974–1987. doi:10.1101/gr.223727.117

Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GAB, Touzet P. 2019. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. *Genome Biol Evol* **11:** 350–361. doi:10.1093/gbe/evz001

Mátés L, Izsvák Z, Ivics Z. 2007. Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. *Genome Biol* **8:** S1. doi:10.1186/gb-2007-8-s1-s1

Matyunina LV, Bowen NJ, McDonald JF. 2008. LTR retrotransposons and the evolution of dosage compensation in *Drosophila*. *BMC Mol Biol* **9:** 55. doi:10.1186/1471-2199-9-55

Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. 2015. The human transcriptome across tissues and individuals. *Science* **348:** 660–665. doi:10.1126/science.aaa0355

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

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol* **10:** e1001308. doi:10.1371/journal.pbio.1001308

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

Nozawa M, Fukuda N, Ikeo K, Gojobori T. 2014. Tissue- and stage-dependent dosage compensation on the neo-X chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol* **31:** 614–624. doi:10.1093/molbev/mst239

Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, New York.

Panta M, Mishra A, Hoque MT, Atallah J. 2021. ClassifyTE: a stacking-based prediction of hierarchical classification of transposable elements. *Bioinformatics* **37:** 2529–2536. doi:10.1093/bioinformatics/btab146

Park Y, Wu H. 2016. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* **32:** 1446–1453. doi:10.1093/bioinformatics/btw026

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for *Xist* in X chromosome inactivation. *Nature* **379:** 131–137. doi:10.1038/379131a0

Pérez-Rico YA, Barillot E, Shkumatava A. 2020. Demarcation of topologically associating domains is uncoupled from enriched CTCF binding in developing zebrafish. *iScience* **23:** 101046. doi:10.1016/j.isci.2020.101046

Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26:** 1641–1650. doi:10.1093/molbev/msp077

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* **26:** 342–350. doi:10.1101/gr.193474.115

Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M, et al. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559:** 392–395. doi:10.1038/s41586-018-0273-1

Ramírez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung HR, Lajoie BR, de Wit E, Zhan Y, de Laat W, et al. 2015. High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in *Drosophila*. *Mol Cell* **60:** 146–162. doi:10.1016/j.molcel.2015.08.024

R Core Team. 2022. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Riggs AD, Pfeifer GP. 1992. X-chromosome inactivation and cell memory. *Trends Genet* **8:** 169–174. doi:10.1016/0168-9525(92)90219-T

Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140. doi:10.1093/bioinformatics/btp616

Rosin LF, Chen D, Chen Y, Lei EP. 2022. Dosage compensation in *Bombyx mori* is achieved by partial repression of both Z chromosomes in males. *Proc Natl Acad Sci* **119:** e2113374119. doi:10.1073/pnas.2113374119

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17:** 155–158. doi:10.1038/s41592-019-0669-3

Sahakyan A, Yang Y, Plath K. 2018. The role of *Xist* in X-chromosome dosage compensation. *Trends Cell Biol* **28:** 999–1013. doi:10.1016/j.tcb.2018.05.005

Saldaña-Meyer R, González-Buendía E, Guerrero G, Narendra V, Bonasio R, Recillas-Targa F, Reinberg D. 2014. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev* **28:** 723–734. doi:10.1101/gad.236869.113

Salvadori S, Coluccia E, Deiana A. 2015. Teleost fish lymphocyte culture. In *Fish cytogenetic techniques* (ed. Ozouf-Costaz C, et al.), Chapter 9. CRC Press, Boca Raton, FL. https://www.taylorfrancis.com/books/9781482211993

Sandkam BA, Almeida P, Darolti I, Furman BLS, van der Bijl W, Morris J, Bourne GR, Breden F, Mank JE. 2021. Extreme Y chromosome polymorphism corresponds to five male reproductive morphs of a freshwater fish. *Nat Ecol Evol* **5:** 939–948. doi:10.1101/2020.08.19.258434

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff J-N, Lesch K-P, et al. 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet* **45:** 567–572. doi:10.1038/ng.2604

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148:** 335–348. doi:10.1016/j.cell.2011.11.058

Schmitz RJ, Lewis ZA, Goll MG. 2019. DNA methylation: shared and divergent features across eukaryotes. *Trends Genet* **35:** 818–827. doi:10.1016/j.tig.2019.07.007

Shen W, Le S, Li Y, Hu F. 2016. Seqkit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11:** e0163962. doi:10.1371/journal.pone.0163962

Shi Y, Seto E, Chang LS, Shenk T. 1991. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67:** 377–388. doi:10.1016/0092-8674(91)90189-6

Shoemaker R, Deng J, Wang W, Zhang K. 2010. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* **20:** 883–889. doi:10.1101/gr.104695.109

Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350:** 978–981. doi:10.1126/science.aad3346

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

Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504:** 465–469. doi:10.1038/nature12719

Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. http://www.repeatmasker.org

Sola L, Iaselli V, Rossi AR, Rasch EM, Monaco PJ. 1992. Cytogenetics of bisexual/unisexual species of *Poecilia*. *Cytogenet Genome Res* **60:** 236–240. doi:10.1159/000133347

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24:** 637–644. doi:10.1093/bioinformatics/btn013

Sumner AT. 1972. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res* **75:** 304–306. doi:10.1016/0014-4827(72)90558-7

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* **375:** 20190347. doi:10.1098/rstb.2019.0347

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 EM, 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

Vicoso B, Bachtrog D. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation. *Chromosome Res* **17:** 585–602. doi:10.1007/s10577-009-9053-y

Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. 2017. YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171:** 1573–1588.e28. doi:10.1016/j.cell.2017.11.008

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

Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. 2010. Redefining CpG islands using hidden Markov models. *Biostatistics* **11:** 499–514. doi:10.1093/biostatistics/kxq005

Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45:** 836–841. doi:10.1038/ng.2649

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356:** eaaj2239. doi:10.1126/science.aaj2239

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8:** 28–36. doi:10.1111/2041-210X.12628

Zhou Q, Ellison CE, Kaiser VB, Alekseyenko AA, Gorchakov AA, Bachtrog D. 2013. The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol* **11:** e1001711. doi:10.1371/journal.pbio.1001711

Received May 25, 2023; accepted in revised form October 20, 2023.

Transposon wave remodeled the epigenomic landscape in the rapid evolution of X-Chromosome dosage compensation

David C.H. Metzger, Imogen Porter, Brendan Mobley, et al.

Genome Res. 2023 33: 1917-1931 originally published online November 21, 2023
Access the most recent version at doi:10.1101/gr.278127.123