Transposon wave remodelled the epigenomic landscape in the rapid evolution of a novel X chromosome dosage compensation mechanism

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

[1]Department of Zoology and Biodiversity Research Centre, University of British Columbia, British Columbia, Canada
[2]Biology Department, Reed College, Portland, Oregon, USA
[3]Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA

* Author for Correspondence: David C.H. Metzger, Department of Zoology and Biodiversity Research Centre, University of British Columbia, Canada, dmetzger@zoology.ubc.ca

Summary

Sex chromosome dosage compensation is a model to understand the coordinated regulation of transcription, however the advanced age of the sex chromosomes in model systems make 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 *de novo* origin of a novel dosage compensation system presents a unique opportunity to discover new mechanisms of gene regulation and their evolutionary origins. By combining a new chromosome-level assembly of *P. picta* with whole-genome bisulfite sequencing and RNA-Seq data, we determine that the binding motif of Yin Yang 1 (YY1) is associated with male hypomethylated regions on the X, but not the autosomes. The male-specific hypomethylation of these motifs offers a putative model for male specific upregulation of genes on the X. 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 de novo origin of a new dosage compensation mechanism.

Keywords: DNA methylation, sex chromosome, genome sequence, dosage compensation, Yin Yang 1, YY1, repetitive element, *Poecilia picta*

Introduction

Some organisms with highly diverged sex chromosomes evolve complete dosage compensation to ameliorate the negative effects of haploinsuffiency.[1,2] These complex mechanisms act across the entirety of the X chromosome and represent a key model to study the coordinated regulation of transcription[3] through the integration of sex-specific genomic and epigenomic processes.[4] Chromosome-wide dosage compensation mechanisms utilize sex-specific chromatin modifying complexes to balance expression between the haploid X and diploid autosomes in males, as well as to equalize expression of X-linked genes between the males and females[5,6] regardless of their dosage sensitivity. This is in stark contrast to many other systems with highly diverged sex chromosomes where gene dose is only compensated for those loci that are dosage sensitive.[7]

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. In eutherian mammals, the long-non coding RNA from *Xist* is expressed early in development from the X inactivation centre and spreads along one X chromosome in female cells,[8,9] ultimately leading to the transcriptional inactivation for most of the genes via the accumulation of epigenomic marks that are transmitted to all daughter cells.[10] In *Drosophila*, dosage compensation of the X chromosome in somatic cells is achieved through the male-specific lethal (MSL) complex, which acetylates histone H4,[11] resulting in hyper-expression of X-linked genes in males.[12] Moreover, substantial insight on how dosage compensation mechanisms are established on neo-sex chromosomes has been gleaned from the spread of existing dosage compensation mechanisms to neo-sex chromosomes in *Drosophila*.[13]

However, our understanding of the evolutionary dynamics and mechanisms of sex chromosome divergence and the evolution of *de novo* dosage compensation mechanisms have thus far been limited due to the fact that the model systems for whole chromosome dosage compensation mechanisms are species with ancient sex chromosomes.[13] Understanding the early stages of chromosome regulation in these models requires extensive extrapolation, and can make it difficult to differentiate cause from consequence.

Recently, the first case of complete sex chromosome dosage compensation in fishes was identified in *Poecilia picta,*[14] and the close relative *P. parae.*[15] *P. picta* and *P. parae* are closely related to the common guppy, *P. reticulata*, and the Endler's guppy, *P. wingei*. These species all share the same sex chromosome yet they exhibit a remarkable diversity of sex chromosome divergence. Although several phylogenetic models have been suggested for the origin of the sex chromosomes and dosage compensation in these species,[16] comparative phylogenetic and genomic methods[15,17,18] suggest that the sex chromosomes diverged, and complete dosage compensation evolved in a short interval between 14.8-18.5mya.[15] Importantly, the sex chromosomes of *P. reticulata* and *P. wingei* lack any evidence of dosage compensation,[14] and karyogram data shows extensive homology between the X and Y relative to the highly degraded Y in *P. picta*.[19] 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 mechanism of dosage compensation from a related species has been co-opted, presenting a unique opportunity to study the early *de novo* evolution of a novel dosage compensation mechanism.

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. To facilitate our analysis, we assembled and annotated a chromosome-level genome sequence of a single female using a combination of PacBio HiFi and HiC sequencing. Using whole genome bisulfite sequencing (WGBS), we identify sex-biased differentially methylated regions (DMRs) on the *P. picta* sex chromosome that are highly conserved among tissues. Using motif analysis of the DMR sequences, we identify enrichment of a putative DNA binding motif for the Yin Yang 1 (YY1) transcription factor. These YY1 motifs are highly enriched on the *P. picta* X chromosome as the result of a recent wave of repetitive element invasion, which is absent in closely related species that lack dosage compensation. We also find that these motifs are hypomethylated on the X in males but not in females, or on the autosomes of either sex. Interestingly, YY1 facilitates *Xist* inactivation of the X chromosome in eutherian mammals[20] and the role of YY1 in *P. picta* appears to be a case of evolutionary convergence. Taken together, our results present compelling support for the

remodeling of the X chromosome epigenomic landscape through a wave of repetitive element insertions carrying YY1 motifs. The male-specific hypomethylation of these motifs offers a putative model for male specific upregulation of genes on the X and the first piece of evidence towards elucidating the rapid *de novo* evolution of a dosage compensation mechanism in *P. picta*.

Results

P. picta genome assembly, annotation and statistics

Our assembly of the *P. picta* genome is 744 Mb in total length, with BUSCO completeness of 97.65% and an N50 of 33,053,486, representing a substantial improvement over previous genomic resources for the species.[21] We recovered 23 chromosomes, numbered based on the syntenic comparison with the *P. reticulata* female genome assembly (Figure 1A),[22] where chromosome 12 is the sex chromosome.[23,24] It is worth noting that this numbering system differs from some previous reports.[19,25] 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* and the short-read Illumina-based female *P. reticulata* genome assembly[22] identified several large inversions and rearrangements (Supplemental Figure 1). In contrast, the *P. picta* genome is highly syntenic with the long-read male *P. reticulata*[26] (Supplemental Figure 2), and *Xiphophorus helleri* (Supplemental Figure 3) genomes, suggesting that the apparent inversions in the female *P. reticulata* genome represent mis-assemblies. 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* (Figure 1B). Linkage analysis of the sex determining region in *P. reticulata* has identified the sex determining locus to the distal end of chromosome 12,[23] 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, particularly in the absence of long-read

sequence data. The *P. picta* X chromosome (31.2Mb) also has ~4Mb at the distal end of the chromosome with very little homology to the X chromosome in *P. reticulata* (26.61Mb).

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

Following the split from *P. reticulata* and *P. wingei* ~18.4mya, the Y chromosome in the common ancestor of *P. picta* and *P. parae* rapidly diverged from the X over a period of 3-4my, leading to widespread gene loss and the rapid evolution of a complete dosage compensation mechanism.[14,15,25] 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.[14,27–29]

The effectiveness of dosage compensation mechanisms can vary between tissues.[30] Previous studies of dosage compensation in *P. picta* and *P. parae* focused on the expression from samples consisting of mostly muscle tissue.[14,15] 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 (Figure 1C; statistics available in Supplemental Table 2). However, the X expression was significantly lower in male gonad tissue compared to the average expression of male and female autosomal and female X genes in gonad. This is consistent with a ubiquitous complete dosage compensation mechanism acting in somatic tissues (Figure 1C, Supplemental Table 2).

The capacity of dosage compensation mechanisms to buffer the effects of gene dosage imbalances can be limited for highly expressed genes.[13,31] 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 quantile, the average expression of sex chromosome genes in males is not significantly lower compared to 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* (Figure 1D, Supplemental Table 4). Sex chromosome genes expressed in the head were significantly lower than autosomal genes for the lowest expression quantile in males, however the same pattern was also observed in females suggesting that this difference is not caused by the efficiency of dosage compensation in genes with low expression.

The P. picta sex chromosome is differentially methylated in somatic tissues.

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[32] and maintenance of X inactivation in mammals,[33] or histone acetylation by the MSL complex in *Drosophila.*[34] We therefore examined male and female muscle, liver, and gonad tissues for sex-specific DNA methylation patterns using WGBS data. Whole genome DNA methylation values were similar in male and female muscle and liver tissue, while testis tissue was more highly methylated compared to ovary or somatic tissue (Supplemental Figure 4). Principal component analysis also revealed that DNA methylation patterns in testis are distinct from other tissues (Supplemental Figure 5). Analysis of DNA methylation values at the chromosome level revealed that CpG islands on the X chromosome are more highly methylated in both males and females compared to autosomal DNA methylation in all tissues (Supplemental Figure 6).

We next examined the WGBS data at the single nucleotide level to identify site specific patterns of sex-biased DNA methylation. Male and female gonads exhibit the greatest number of differentially methylated loci (DMLs) (744,852) and 79% of the DMLs are hypo-methylated in testis compared to ovary. DMLs between testis and ovary are evenly distributed throughout the genome (Supplemental Figure 7), with no enrichment on the X chromosome. We found relatively fewer sex-biased DNA methylation sites between male and female somatic tissues (9,575 DMLs in muscle; 7,536 DMLs in liver). In contrast to gonad tissue, sex-specific DNA methylation patterns in somatic tissue were highly enriched on the X chromosome (74% of DMLs in muscle and 82% of DMLs in liver compared to 5% in gonad).

The spatial distribution of CpG sites is not uniform throughout the genome. DMLs tend to occur in clusters forming regions of contiguous differential methylation profiles called Differentially Methylated Regions (DMRs). We therefore next analyzed the WGBS data for sex-biased DMRs. We found that the genomic distribution of DMRs is consistent with the distribution of DMLs (Figure 2, Supplemental Figure7-8). We found 1268 sex-biased DMRs between gonad tissues, 295 DMRs in muscle tissue, and 221 DMRs in liver tissue. The majority of DMRs in somatic tissues are hypomethylated in males (80% in muscle and 84% in liver) and are enriched on the sex chromosome (20% in muscle and 63% in liver), while the distribution of sex-specific DMRs in gonad were equally hypo- and hypermethylated in gonad tissue (53% hypomethylated and 47% hypermethylated in males), and were evenly distributed among autosomes and the sex chromosome (Figure 2, Supplemental Figure 7).

Hypomethylated DMRs in promoter regions can activate the expression of nearby genes resulting in sex biased expression.[35] Consistent with having a role in regulating gene expression, we found lower levels of DNA methylation in the promoter regions of highly expressed genes (Supplemental Figure 9). To determine whether DMRs on the sex chromosome are involve in gene-by-gene dosage compensation, we next examined DNA methylation levels of CpG islands located in the promoter regions of genes. We found no evidence for chromosome-wide sex-biased DNA methylation of gene promoters (Supplemental Figure 10), 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.

Some cases of sex chromosome dosage compensation are associated with heterochromatization.[36,38] We therefore examined male and female *P. picta* karyograms for evidence of chromosome wide differences in heterochromatinization between male and female X chromosomes. Consistent with with karyograms of *P. picta* from a Trinidadian population (Nanda et al., 2022), we identified 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 Figure 11) suggesting that overall heterochromatization 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 ten 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 non-coding RNA and dosage compensation in mammals.[20,37,39–41]

Further analysis of the YY1 sequence motif revealed a dramatic abundance of YY1 motifs on the sex chromosome compared to the autosomes, likely resulting from the replication of three of the six possible motifs (CCGCCATATTGT, CGGCCATTTTGT, and CAGCCATTTTGT) on the sex chromosome relative to the autosomes (Figure 3A). These motifs are highly conserved compared to 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.[42]

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 (Figure 3A). YY1 sequence motifs were enriched on the X chromosome of the sister species *P. parae*, which also exhibits complete X chromosome dosage compensation.[15] We found no enrichment of YY1 motifs on the X chromosome of *P. reticulata*, which shares the same X chromosome but lacks substantial Y degeneration or dosage compensation[14] or *P. latippina,* which has a ZW sex determining mechanism located on a different chromosome from *P. picta* [14,43] (Figure 3A).

Comparison of the YY1 motif distribution along the X chromosome revealed a striking similarity to the distribution of DMRs, particularly those in somatic (i.e. dosage compensated) tissues (Figure 3B). YY1 motifs have been shown to be regulated by differential DNA methylation patterns in other species[37,44] and is known to play a role in *Xist* mediated X chromosome inactivation in eutherian mammals.[20,40,45] 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 (Figure 3C). The percent difference in DNA methylation of YY1 motifs on the sex chromosome were similar in muscle (50%) and liver (48%) tissue. DNA methylation of YY1 motifs in testis were also hypomethylated but to a lesser extent (29%) compared to somatic tissues. In contrast, we observe no difference between male and female DNA methylation levels for YY1 motifs on the autosomes (Figure 3C).

To determine whether male-specific hypomethylation of YY1 motifs results in a localized hyper-expression of genes, we also analyzed gene expression levels in the immediate 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 Figure 12). These data demonstrate that methylation of YY1 sites does not regulate expression on a gene-by-gene basis and suggests a putative dosage compensation mechanism mediated by YY1 binding affinity and long-range control of gene expression. Annotation of the *P. picta* genome identified two YY1 orthologs, ANN22864 on ch21 and ANN18583 on chr22. Both YY1 orthologs are ubiquitously expressed in head, muscle, liver, and gonad tissues (Supplemental Figure 13) and their DNA binding domain, consisting of four C2H2-type zinc fingers, is identical to the mammalian YY1 DNA binding domain and the DNA binding domain of other fishes (Supplemental Figure 14).

***Recent expansion of repetitive elements containing YY1 motifs*.**

YY1 motifs can regulate the expression of LINE-1 transposable elements (TE),[46,47] 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 (Figure 3D). This phylogenetic analysis revealed that most of these unannotated repetitive sequences cluster with sequences from across the autosomes, but three distinct clusters contain repetitive sequences that are present, almost exclusively, on the X chromosome. 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 200kb sliding windows for each chromosome. Of the 23 chromosomes, the sex

chromosome contains an anomalous abundance of longer repeats (Figure 3E) with low Kimura distance values (Supplemental Figure 15) consistent with a recent expansion of repetitive elements on this chromosome, which may partially account for the size discrepancy between the X chromosome assembly size between *P. picta* (31.3Mb) and *P. reticulata* (26.6Mb) (Figure 1B). Taken together, these results suggest that a recent expansion of transposable elements along the X chromosome in *P. picta* produced the rapid accumulation of the YY1 transcription factor motifs.

Discussion

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.[3,48] However, it remains unclear how these complex adaptations to Y chromosome degeneration can arise, as the sex chromosomes of model systems are ancient, making it difficult to extrapolate the initial stages of the *de novo* origin of dosage compensation, and obscuring cause from consequence. 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 mechanism. Our results support the model that a recent wave of transposon replication led to a sex-chromosome specific accumulation of YY1 DNA binding motifs and that sex-biased DNA methylation promotes male specific activation of the YY1 motifs as a key component in establishing dosage compensation in *P. picta* (Figure 4).

YY1 and dosage compensation in Poecilia picta

YY1 is a ubiquitously expressed[49,50] DNA-binding zinc-finger transcription factor [51,52] that has been shown to selectively bind to hypo-methylated DNA sequences.[53,54] In mammals, YY1 shares several characteristics with the conserved CCCTC-binding factor protein (CTCF). Both CTCF and YY1 are ubiquitously expressed transcription factors that function as activators, repressors, chromatin insulators, and play important roles in imprinting.[41] CTCF protein has a well-established role in gene regulation and the formation of 3D chromatin organization such as

topologically associated domains in mammals.[55] These gene expression domains are often demarcated by the formation of newly derived CTCF motifs resulting from a recent expansion of repetitive elements.[56] Analogous to CTCF-CTCF interactions, YY1 dimerization has also been shown to promote the formation of DNA loops, forming insulated regions of and facilitating enhancer-promoter interactions.[39,49,57,58]

Our results are consistent with a recent expansion of YY1 motifs and the compartmentalization of the X chromosome in the common ancestor of *P. picta* and *P. parae.* These data suggest a convergent model for the evolution of regulatory domains and novel dosage compensation mechanisms through a transposable element mediated expansion of transcription factor motifs.[59,60] Additionally, allele specific expression patterns in *P. picta* and *P. parae* strongly suggest that dosage compensation in these species is achieved through a male-specific hyperexpression of the X chromosome,[14,15] a process comparable to dosage compensation in *Drosophila.*[34] Our study suggests that male-specific hyperexpression of the X chromosome is achieved by male-specific hypomethylation of YY1 motifs on the X chromosome, thereby facilitating male-specific activation of these motifs and modifying expression of the entirety of the X chromosome. Importantly, transcription of YY1 regulatory elements produces a positive-feedback loop that contributes to the stability of gene expression programs [61] which could provide a mechanism to maintain sex-specific activity of YY1 motifs once they are established.

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 non-coding RNA involved X chromosome silencing in mammals.[20,37,39–41] In the context of mammalian dosage compensation, YY1 binding is differentially regulated by the sex-specific epigenetic landscape between the active and inactive X in females. Our data supports the convergent evolution of a YY1 mediated dosage compensation mechanism regulated by the epigenomic landscape in *P. picta*. While the exact role of YY1 in *P. picta* is still unclear, future studies will provide key insights into the role of YY1 and the regulation of chromosome structure, enhancer-promoter interactions, and as a nucleation factor for the binding and spreading of dosage compensation machinery[20,39,41] on the *P. picta* X chromosome. Although some cases of sex chromosome dosage compensation are associated with heterochromatization,[36,38] morphological analysis of metaphase spreads from

368 males and females in *P. picta* showed no clear difference between male and female X
369 chromosome morphology. Comparison of male and female X chromosomes using higher
370 resolution methods, such as C-banding which can selectively stain heterochromatin-rich regions
371 of individual chromosomes, may be more effective at identifying sex-specific differences in X
372 chromosome chromatin organization[62,63], or it may be that X chromosome dosage
373 compensation is achieved without differential heterochromatinization between males and
374 females, as is the case in *Drosophila*.[67]

375 *TE replication and sex chromosome evolution*

376 Transposable elements are major innovators of genome evolution and speciation. Replication
377 and expansion of genomic regions via transposable element activity is an important process
378 that contributes to the generation of genomic diversity such as the replication and formation of
379 new TF binding sites.[56,64,65] In some cases, notably the case of the *Drosophila* neo sex
380 chromosomes,[66] existing mechanisms of dosage compensation can spread with the expansion
381 of binding motifs.[67] However, many of the genes involved in the formation and spreading of the
382 *Drosophila* MSL complex are not present in fish genomes.

383 Younger, potentially active TEs are silenced by epigenetic processes such as DNA
384 methylation.[68,69] In particular, LINE-1 retrotransposition is regulated by DNA methylation of YY1
385 transcription factor motifs.[46,47,70] Furthermore, the copy and paste mechanism through which
386 the expansion of retrotransposons occurs means that once a YY1 motif is established in a TE, it
387 has the potential to spread throughout the genome. This provides a potential positive feedback
388 loop for the expansion of TEs bearing YY1-motifs in hypomethylated regions.

389 We have identified a recent repeat-associated expansion YY1-binding motifs on the *P. picta* X
390 chromosome, suggesting a potential model for the rapid evolution of dosage compensation.
391 The X-chromosome enrichment of YY1 motifs and YY1-bearing TEs is absent from the syntenic
392 chromosomes of *P. reticulata* and *P. latippina*, which both lack dosage compensation. This
393 phylogenetic distribution suggests that the expansion occurred in the common ancestor to *P.*
394 *picta* and *P. parae* ~18.4mya in tandem with the rapid evolution of dosage compensation over
395 3-4myr[15] (Figure 3A). 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*.[14,16,18] However, all available evidence suggests that the sex chromosome arose on in the ancestor to *P. reticulata*, *P. wingei*, *P. picta* and *P. parae*, roughly 20 mya.[15,17,18,25] 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, roughly 20mya. 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. Because Y chromosome degeneration is facilitated by the evolution of dosage compensation[71] 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. 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,[16] 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 model of a recent transposable element wave that resulted in the accumulation of YY1 elements on the X, again consistent with a recent origin of dosage compensation.

Conclusion

Understanding the structure, function, and origins of genomic processes is fundamental to understanding the mechanisms of genomic evolution. *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,[72] however this mechanism remains unknown. The extensive phylogenetic distance to other known mechanisms of whole chromosome dosage compensation make 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. 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 transposable element expansion coupled with sex-specific regulation of the YY1 regulatory motif via DNA methylation. These data support the convergent evolution of YY1 as key component of dosage compensation between mammals and Poeciliids. However, the putative dosage compensation mechanisms described here represents a novel function of YY1, as fishes lack many of the genes and regulatory systems involved in mammalian and *Drosophila* dosage compensation mechanisms.

Materials and 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 2019 from natural populations in Suriname and brought to the University of British Columbia (Vancouver, BC, Canada) aquatics facility, where they were kept in 20 L glass aquaria on a 12:12 day:night cycle at 26 °C and 13ppt salinity (Instant Ocean Sea Salt), and fed Hikari Fancy Guppy pellets supplemented with live brine shrimp daily. Prior to 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 ^{o}C.

PacBio Library and Sequencing

Frozen tissue samples were shipped overnight on dry ice to the Dovetail Genomics Production Lab (Scotts Valley, CA) for Dovetail Omni-C Library Preparation and Sequencing. DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using the manufacturer 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[73] 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[74] 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 were identified as contaminant and were removed from the assembly. The filtered assembly was then used as an input to purge_dups v1.1.2[75] 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 HiSeqX platform to produce approximately 40x sequence coverage. Reads with MQ>50 were then used for scaffolding.

Scaffolding the Assembly with HiRise

The input de novo assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies.[76] Dovetail OmniC library sequences were aligned to the draft input assembly using bwa.[77] The separations of Dovetail OmniC 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 Benchmarking Universal Single-Copy Orthologs (BUSCOs) against the eukaryota_odb10 database.[78]

Genome annotation

Coding sequences from *P. formosa*, *P. latipinna*, *P. 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 optimisation were done with the software package provided by AUGUSTUS. The same coding sequences were also used to train a separate ab initio model for *P. picta* using SNAP (version 2006-07-28). RNAseq reads were mapped onto the genome using the STAR aligner software (version 2.7) and intron hints generated with the bam2hints tools within the AUGUSTUS software. MAKER, SNAP 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 softwares 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 characterised 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 (version 2.05).

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 250bp PE

reads. Reads were trimmed using Trimmomatic[79] resulting in 1.2 million sequences, which were de novo assembled using Geneious Prime® (v.2022.2.1 Biomatters Ltd.). The resulting mitochondrial sequence was aligned sequences of close relatives (*P. parae*, *P. reticulata* and *P. formosa*) and annotations were lifted over. A 245bp insertion in the D-Loop of *P. picta* was confirmed using 3 sanger reactions to sequence the 1.7kb region spanning the insertion.

Synteny analysis

Sequence synteny was compared between the *P. picta* genome and the genomes of *P. reticulata* (female Genbank accession# GCA_000633615.2 and male Genbank accession # GCA_904066995.1) and *X. hellerii* (Genbank accession# GCA_003331165.2). Syntenic regions were visualized using the output from satsuma2 in Circos.[80] Synteny between the *P. picta* and *P. reticulata* X chromosome was visualized using plotsr.[81] 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 sequencers was performed in Geneious (v.2022.2.1 Biomatters Ltd.).

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.[82] Briefly, live specimens were exposed to a 0.01% colchicine solution for seven hours before being euthanized with a lethal overdose of MS-222 and dissected. Extracted organs were incubated in 0.4% KCl solution for 25-45 minutes 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 approximately 45°C, then stained for 10 minutes 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 (100x/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.[83] All chromosomes prepared from female specimens were successfully paired while 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 (i.e., pair 1).

RNA-seq sampling and analysis

Sampling of adult fish was performed as described above. Liver, head, and gonad tissue was sampled for RNA-seq analysis and genome annotation. 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 non-overlapping male and three female sample pools for library preparation and sequencing. Frozen tissue samples were shipped on dry ice to Genewiz (South Plainfield, NJ) for RNA purification, mRNA library preparation and sequencing.

Total RNA extraction was done using the QIAGEN RNeasy Plus Kit following manufacturer protocols. Total RNA was quantified using Qubit RNA Assay and TapeStation 4200. Prior to library prep, RNA samples were DNase treated followed by AMPure bead clean up and QIAGEN FastSelect HMR rRNA depletion. Library preparation was done with the NEBNext Ultra II RNA Library Prep Kit following manufacturer protocols. Libraries were then sequenced on the NovaSeq6000 platform in 2 x 150 bp configuration. RNA-seq data from muscle tissue (SRA accession #'s SAMN29631966, SAMN29631977, SAMN29631978, SAMN29631989, SAMN31093659)[18] was also used for genome annotation and RNA-seq analysis.

563 Sequencing data was quality filtered and Illumina adapters were trimmed using Trimmomatic
564 v0.36.[79] Reads were scanned with a 4-base sliding window and discarded when the average
565 phred score was <15. Reads were trimmed to 120bps and the first 10bps were trimmed from
566 each read. Reads were mapped to the *P. picta* reference genome using hisat2 v2.1.[84] Only
567 uniquely mapped reads that aligned as pairs were retained for further analysis. Average
568 mapping efficiency was 97.7% with an average mapped library size of 67,004,220 reads
569 (Supplemental Table 2). Alignment bam files were sorted and indexed using samtools v1.3.1.[85]
570 Aligned reads were assigned to genic regions using featureCounts v2.0.1.[86]

571 Gene expression analysis was conducted in R v4.1 using edgeR v3.34.1.[87] Genes with zero
572 counts in all samples were removed. We then filtered the data to retain only genes with >10
573 reads in each female or male sample. This filtered dataset was then used to calculate
574 normalization factors using the RLE method in the calcNormFactors() function. We estimated
575 robust dispersions using the estimateDisp() followed by the estimateGLMRobustDisp() function.
576 Normalized counts per million (cpm) values were then extracted from the data to compare
577 gene expression levels between males and females and between genes located on the X
578 chromosome compared to the autosomes. For the quantile expression analysis, the quantile()
579 function from the R stats package was used to bin genes into the 0-25, 25-50, 50-75, and 75-
580 100th percentiles based on gene expression values in males.[31] A Wilcoxon rank sum test was
581 used to identify differences in gene expression levels between groups using an FDR corrected p-
582 value <0.05.

583

584 *Bisulfite Sequencing sampling and analysis*

585 Muscle, liver, and gonad tissues were sampled from adult male and virgin female *P. picta* as
586 described above. A total of 15 males and 15 females were sampled. To minimize the effect of
587 genetic differences and developmental factors contributing to inter-individual variation in DNA
588 methylation patterns,[88] we pooled tissues from five male or five female tissue samples to
589 generate a total of three male and three female non-overlapping sample pools for DNA
590 isolation and sequencing with the exception of ovary tissue where we were only able to extract

enough DNA from 2 of the 3 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.[79] Reads were scanned with a 4-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.[89] 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 is 105,671,545 reads (Supplemental Table 5). BAM alignment files were prepared for duplicate removal using samtools v1.15.1 fixmate followed by samtools sort.[85] 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.[90] Differentially methylated loci (DMLs) and differentially methylated regions (DMRs) between males and females were identified for each tissue using the dmlTest() function with smoothing = TRUE followed by callDML() and callDMR() respectively. CpG islands were identified using a hidden markov model using the makeCGI_1.3.4 R package.[91]

Motif enrichment and YY1 analysis

Motif enrichment analysis for differentially methylated regions on the X chromosome was conducted using findMotifsGenome.pl in Homer v4.11.1.[92] 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 v2.60.2 R package to search for YY1 motifs in the *P. picta* genome and to also identify repetitive

regions containing YY1 motifs. DNA methylation of YY1 motifs was calculated as the mean DNA methylation value of CpG loci +/- 500bps 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[93] to generate a repeat element library. We then used RepeatMasker V4.1.2[94] 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.[95] The output file from the Muscle alignment was then processed using FastTree V2.1.[96] The phylogeny from FastTree was visualized using the ggtree v3.0.4 package in R. To compare TE length and abundance between chromosomes, we calculated the total length (bp) and number of TEs in 200kb non-overlapping sliding windows. The mean total length and count per 200kb for each scaffold were then plotted against each other in a scatter plot with standard error bars and a linear regression line.

Data Availability

The *Poecilia picta* genome assembly is available from the NCBI Genome database BioProject PRJNA862953. Raw RNAseq sequence files are available from the NCBI sequence read archive (SRA) BioProject PRJNA884377 (SRA accession #'s SAMN31026551 - SAMN31026568). Raw whole genome bisulfite sequence files are available from the SRA Bioproject PRJNA884372 (SRA accession #'s SAMN31025781 - SAMN31025797). Computer code used for data processing and analysis are available at https://github.com/manklab/Metzger_et_al_picta_DC.

Acknowledgements

We would like to thank Jacelyn Shu for the scientific illustrations in Figure 4, for help with animal husbandry, and tissue dissections for karyotype analysis. We thank Clara Lacy for the drawings of male and female *Poecilia picta*. We would also like to thank Wouter van der Bijl for assistance with R and members of the Mank lab for helpful and constructive suggestions on this

manuscript. This work was funded by grants from the European Research Council (grant number 680951), NSERC and CFI, as well as a Canada 150 Research Chair to JEM. We also acknowledge funding from a Dovetail Genomics Tree of Life Award to DCHM and funding from NSERC Undergraduate Student Research Award grant awarded to IP.

Author Contributions

DCHM and JEM conceived the study and wrote the manuscript. DCHM performed tissue collection and sample preparation for genome sequencing, WGBS, and RNAseq data, performed bioinformatic analysis of WGBS and RNAseq data, motif enrichment, and characterization of YY1 repetitive elements. IP identified and characterized repetitive elements in the *P. picta* genome. BM and AA performed the karyotype analysis. BAS sequenced the mitochondrial genome, performed YY1 amino acid alignments, and helped with sample collection. LJMV helped with sample collection.

Declaration of interests

The authors declare no competing interests

References

1. Ohno, S. (1967). Sex chromosomes and sex-linked genes Ohno, ed. (Springer-Verlag).
2. Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet *14*, 113–124. 10.1038/nrg3366.
3. Kuroda, M.I., Hilfiker, A., and Lucchesi, J.C. (2016). Dosage Compensation in Drosophila —a Model for the Coordinate Regulation of Transcription. Genetics *204*, 435–450. 10.1534/genetics.115.185108.
4. Ferrari, F., Alekseyenko, A.A., Park, P.J., and Kuroda, M.I. (2014). Transcriptional control of a whole chromosome: emerging models for dosage compensation. Nat Struct Mol Biol *21*, 118–125. 10.1038/nsmb.2763.
5. Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet *7*, 645–653. 10.1038/nrg1914.
6. Mank, J.E. (2009). The W, X, Y and Z of sex-chromosome dosage compensation. Trends in Genetics *25*, 226–233. 10.1016/j.tig.2009.03.005.
7. Mank, J.E. (2013). Sex chromosome dosage compensation: definitely not for everyone. Trends in Genetics *29*, 677–683. 10.1016/j.tig.2013.07.005.

8. Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469. 10.1038/nature12719.
9. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137. 10.1038/379131a0.
10. Sahakyan, A., Yang, Y., and Plath, K. (2018). The Role of Xist in X-Chromosome Dosage Compensation. Trends Cell Biol *28*, 999–1013. 10.1016/j.tcb.2018.05.005.
11. Turner, B.M., Birley, A.J., and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. Cell *69*, 375–384. 10.1016/0092-8674(92)90417-B.
12. Bone, J.R., Lavender, J., Richman, R., Palmer, M.J., Turner, B.M., and Kuroda, M.I. (1994). Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. Genes Dev *8*, 96–104. 10.1101/gad.8.1.96.
13. Vicoso, B., and Bachtrog, D. (2009). Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosome Research *17*, 585–602. 10.1007/s10577-009-9053-y.
14. Darolti, I., Wright, A.E., Sandkam, B.A., Morris, J., Bloch, N.I., Farré, M., Fuller, R.C., Bourne, G.R., Larkin, D.M., Breden, F., et al. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. Proceedings of the National Academy of Sciences *116*, 19031–19036. 10.1073/pnas.1905298116.
15. Metzger, D.C.H., Sandkam, B.A., Darolti, I., and Mank, J.E. (2021). Rapid evolution of complete dosage compensation in *Poecilia*. Genome Biol Evol *13*, 1–7. 10.1093/gbe/evab155.
16. Charlesworth, D., Bergero, R., Graham, C., Gardner, J., and Keegan, K. (2021). How did the guppy Y chromosome evolve? PLoS Genet *17*, e1009704. 10.1371/journal.pgen.1009704.
17. Darolti, I., Fong, L.J.M., and Mank, J.E. (2021). Sex chromosome heteromorphism and the fast-X effect in Poeciliids. 1–25.
18. Fong, L.J.M., Darolti, I., Metzger, D.C., Morris, J., Lin, Y., Sandkam, B.A., and Mank, J.E. Evolutionary history of the *Poecilia picta* sex chromosomes. in review, 5.
19. Nanda, I., Schories, S., Simeonov, I., Adolfi, M.C., Du, K., Steinlein, C., Alsheimer, M., Haaf, T., and Schartl, M. (2022). Evolution of the degenerated Y-chromosome of the swamp guppy, *Micropoecilia picta*. Cells *11*, 1118. 10.3390/cells11071118.
20. Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133. 10.1016/j.cell.2011.06.026.
21. Charlesworth, D., Graham, C., Trivedi, U., Gardner, J., and Bergero, R. (2021). PromethION sequencing and assembly of the genome of *Micropoecilia picta*, a fish with a highly degenerated Y chromosome. Genome Biol Evol *13*, 1–11. 10.1093/gbe/evab171.
22. Künstner, A., Hoffmann, M., Fraser, B.A., Kottler, V.A., Sharma, E., Weigel, D., and Dreyer, C. (2016). The genome of the Trinidadian Guppy, *Poecilia reticulata*, and variation in the Guanapo population. PLoS One *11*, e0169087. 10.1371/journal.pone.0169087.
23. Tripathi, N., Hoffmann, M., Weigel, D., and Dreyer, C. (2009). 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. 10.1534/genetics.108.098541.

24. Tripathi, N., Hoffmann, M., Willing, E.M., Lanz, C., Weigel, D., and Dreyer, C. (2009). Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. Proceedings of the Royal Society B: Biological Sciences *276*, 2195–2208. 10.1098/rspb.2008.1930.

25. Sandkam, B.A., Almeida, P., Darolti, I., Furman, B., van der Bijl, W., Morris, J., Bourne, G.R., Breden, F., and Mank, J.E. (2021). Extreme Y chromosome polymorphism corresponds to five male reproductive morphs. Nat Ecol Evol. 10.1101/2020.08.19.258434.

26. Fraser, B.A., Whiting, J.R., Paris, J.R., Weadick, C.J., Parsons, P.J., Charlesworth, D., Bergero, R., Bemm, F., Hoffmann, M., Kottler, V.A., et al. (2020). Improved reference genome uncovers novel sex-linked regions in the guppy (*Poecilia reticulata*). Genome Biol Evol *12*, 1789–1805. 10.1093/gbe/evaa187.

27. Wright, A.E., Darolti, I., Bloch, N.I., Oostra, V., Sandkam, B., Buechel, S.D., Kolm, N., Breden, F., Vicoso, B., and Mank, J.E. (2017). Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nat Commun *8*, 1–10. 10.1038/ncomms14251.

28. Almeida, P., Sandkam, B.A., Morris, J., Darolti, I., Breden, F., and Mank, J.E. (2020). Divergence and remarkable diversity of the Y chromosome in guppies. Mol Biol Evol, 2020.07.13.200196. 10.1093/molbev/msaa257.

29. Darolti, I., Wright, A.E., and Mank, J.E. (2020). Guppy Y chromosome integrity maintained by incomplete recombination suppression. Genome Biol Evol *12*, 965–977. 10.1093/gbe/evaa099.

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

31. Harrison, P.W., Mank, J.E., and 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. 10.1093/gbe/evs086.

32. Golden, L.C., Itoh, Y., Itoh, N., Iyengar, S., Coit, P., Salama, Y., Arnold, A.P., Sawalha, A.H., and Voskuhl, R.R. (2019). Parent-of-origin differences in DNA methylation of X chromosome genes in T lymphocytes. Proc Natl Acad Sci U S A *116*, 26779–26787. 10.1073/pnas.1910072116.

33. Riggs, A.D., and Pfeifer, G.P. (1992). X-chromosome inactivation and cell memory. Trends in Genetics *8*, 169–174. 10.1016/0168-9525(92)90219-T.

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

35. Schmitz, R.J., Lewis, Z.A., and Goll, M.G. (2019). DNA Methylation: Shared and Divergent Features across Eukaryotes. Trends in Genetics *35*, 818–827. 10.1016/j.tig.2019.07.007.

36. Chadwick, B.P., and Willard, H.F. (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. Proc Natl Acad Sci U S A *101*, 17450–17455. 10.1073/pnas.0408021101.

37. Kim, J. do, Hinz, A.K., Bergmann, A., Huang, J.M., Ovcharenko, I., Stubbs, L., and Kim, J. (2006). Identification of clustered YY1 binding sites in imprinting control regions. Genome Res *16*, 901–911. 10.1101/gr.5091406.
38. Rosin, L.F., Chen, D., Chen, Y., and Lei, E.P. (2022). Dosage compensation in Bombyx mori is achieved by partial repression of both Z chromosomes in males. Proc Natl Acad Sci U S A *119*, 1–12. 10.1073/pnas.2113374119.
39. Chen, C.Y., Shi, W., Balaton, B.P., Matthews, A.M., Li, Y., Arenillas, D.J., 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*, 1–14. 10.1038/srep37324.
40. Makhlouf, M., Ouimette, J.F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun *5*, 1–12. 10.1038/ncomms5878.
41. Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Mol Cell *25*, 43–56. 10.1016/j.molcel.2006.11.017.
42. Kim, J. do, and Kim, J. (2009). YY1's longer DNA-binding motifs. Genomics *93*, 152–158. 10.1016/j.ygeno.2008.09.013.
43. Sola, L., Iaselli, V., Rossi, A.R., Rasch, E.M., and Monaco, PJ. (1992). Cytogenetics of bisexual/unisexual species of *Poecilia*. Cytogenet Genome Res *60*, 236–240. 10.1159/000133347.
44. Cusack, M., King, H.W., Spingardi, P., Kessler, B.M., Klose, R.J., and Kriaucionis, S. (2020). Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. Genome Res *30*, 1393–1406. 10.1101/GR.257576.119.
45. Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet *15*, 89. 10.1186/s12863-014-0089-4.
46. Becker, K.G., Swergold, G., Ozato, K., and Thayer, R.E. (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. 10.1093/hmg/2.10.1697.
47. Athanikar, J.N., Badge, R.M., and Moran, J. v. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res *32*, 3846–3855. 10.1093/nar/gkh698.
48. Dekker, J., and Mirny, L. (2016). The 3D Genome as moderator of chromosomal communication. Cell *164*, 1110–1121. 10.1016/j.cell.2016.02.007.
49. Weintraub, A.S., Li, C.H., Zamudio, A. v, Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a structural regulator of enhancer-promoter loops. Cell *171*, 1573-1588.e28. 10.1016/j.cell.2017.11.008.
50. Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). The human transcriptome across tissues and individuals. Science (1979) *348*, 660–665. 10.1126/science.aaa0355.
51. Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V. v (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. 10.1128/mcb.13.12.7612-7624.1993.

52. Shi, Y., Seto, E., Chang, L.S., and 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. 10.1016/0092-8674(91)90189-6.

53. Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485. 10.1038/35013100.

54. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (1979) *356*. 10.1126/science.aaj2239.

55. Pérez-Rico, Y.A., Barillot, E., and Shkumatava, A. (2020). Demarcation of topologically associating domains is uncoupled from enriched CTCF binding in developing Zebrafish. iScience *23*, 101046. 10.1016/j.isci.2020.101046.

56. Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonalves, Â., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348. 10.1016/j.cell.2011.11.058.

57. López-Perrote, A., Alatwi, H.E., Torreira, E., Ismail, A., Ayora, S., Downs, J.A., and Llorca, O. (2014). Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. Journal of Biological Chemistry *289*, 22614–22629. 10.1074/jbc.M114.567040.

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

59. Ellison, C.E., and Bachtrog, D. (2013). Dosage compensation via transposable element mediated rewiring of a regulatory network. Science (1979) *342*, 846–850. 10.1126/science.1239552.

60. Matyunina, L. v., Bowen, N.J., and McDonald, J.F. (2008). LTR retrotransposons and the evolution of dosage compensation in *Drosophila*. BMC Mol Biol *9*, 1–8. 10.1186/1471-2199-9-55.

61. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science (1979) *350*, 978–981. 10.1126/science.aad3346.

62. Salvadori, S., Coluccia, E., and Deiana, A. (2015). Fish Cytogenetic Techniques C. Ozouf-Costaz, E. Pisano, F. Foresti, and L. F. de A. Toledo, eds. (CRC Press) 10.1201/b18534.

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

64. Sundaram, V., and Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philosophical Transactions of the Royal Society B: Biological Sciences *375*. 10.1098/rstb.2019.0347.

65. Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., et al. (2013). DNA hypomethylation within specific transposable

element families associates with tissue-specific enhancer landscape. Nat Genet *45*, 836–841. 10.1038/ng.2649.

66. Marín, I., Franke, A., Bashaw, G.J., and Baker, B.S. (1996). The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. Nature *383*, 160–163. 10.1038/383160a0.

67. Zhou, Q., Ellison, C.E., Kaiser, V.B., Alekseyenko, A.A., Gorchakov, A.A., and Bachtrog, D. (2013). The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. PLoS Biol *11*, e1001711. 10.1371/journal.pbio.1001711.

68. Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96–99. 10.1038/nature02886.

69. Nagamori, I., Kobayashi, H., Shiromoto, Y., Nishimura, T., Kuramochi-Miyagawa, S., Kono, T., and Nakano, T. (2015). Comprehensive DNA methylation analysis of retrotransposons in male germ cells. Cell Rep *12*, 1541–1547. 10.1016/j.celrep.2015.07.060.

70. Sanchez-Luque, F.J., Kempen, M.-J.H.C., Gerdes, P., Vargas-Landin, D.B., Richardson, S.R., Troskie, R.-L., Jesuadian, J.S., Cheetham, S.W., Carreira, P.E., Salvador-Palomeque, C., et al. (2019). LINE-1 evasion of epigenetic repression in humans. Mol Cell *75*, 590-604.e12. 10.1016/j.molcel.2019.05.024.

71. Lenormand, T., and Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. Science (1979) *375*, 663–666. 10.1126/science.abj1813.

72. Marin, R., Cortez, D., Lamanna, F., Pradeepa, M.M., 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. 10.1101/gr.223727.117.

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

74. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: interrogation of genome assemblies. F1000Res *6*, 1287. 10.12688/f1000research.12232.1.

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

76. Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res *26*, 342–350. 10.1101/gr.193474.115.

77. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *00*, 1–3. 10.48550/arXiv.1303.3997.

78. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. v., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. 10.1093/bioinformatics/btv351.

79. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. 10.1093/bioinformatics/btu170.

80. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res *19*, 1639–1645. 10.1101/gr.092759.109.
81. Goel, M., and Schneeberger, K. (2022). plotsr: Visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics *38*, 2922–2926. 10.1093/bioinformatics/btac196.
82. Kligerman, A.D., and Bloom, S.E. (1977). Rapid chromosome preparations from solid tissues of fishes. Journal of the Fisheries Research Board of Canada *34*, 266–269. 10.1139/f77-039.
83. Levan, A., Fredga, K., and Sandberg, A.A. (2009). Nomenclature for centromeric chromosome position on chromosomes. Hereditas *52*, 201–220. 10.1111/j.1601-5223.1964.tb01953.x.
84. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907–915. 10.1038/s41587-019-0201-4.
85. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. 10.1093/bioinformatics/btp352.
86. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. 10.1093/bioinformatics/btt656.
87. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. 10.1093/bioinformatics/btp616.
88. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33*, 245–254. 10.1038/ng1089.
89. Farrell, C., Thompson, M., Tosevska, A., Oyetunde, A., and Pellegrini, M. (2021). BiSulfite Bolt: A bisulfite sequencing analysis platform. Gigascience *10*, 1–8. 10.1093/gigascience/giab033.
90. Park, Y., and Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics *32*, 1446–1453. 10.1093/bioinformatics/btw026.
91. Wu, H., Caffo, B., Jaffee, H.A., Irizarry, R.A., and Feinberg, A.P. (2010). Redefining CpG islands using hidden Markov models. Biostatistics *11*, 499–514. 10.1093/biostatistics/kxq005.
92. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589. 10.1016/j.molcel.2010.05.004.
93. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A *117*, 9451–9457. 10.1073/pnas.1921046117.
94. Smit, A.F., Hubley, R., and Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org.

95. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113. 10.1186/1471-2105-5-113.
96. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol *26*, 1641–1650. 10.1093/molbev/msp077.
97. Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M., et al. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature *559*, 392–395. 10.1038/s41586-018-0273-1.

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

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

Figure Legends

Figure 1. *Poecilia picta* genome assembly and analysis of tissue specific dosage compensation. A. Sequence synteny between the female *P. picta* and the female *P. reticulata* genomes. Chromosomes are labeled 1-23 and are differentiated by colour around the outer ring with *P. picta* chromosomes on the left and *P. reticulata* chromosomes on the right. The colour of the connecting lines indicates the chromosomal origin of the sequence from the *P. picta* genome. B. Chromosome 12 alignments between the female *P. reticulata*[22] (red horizontal line), our *P. picta* (green horizontal line), and the male *P. reticulata*[26] (blue horizontal line) genome assemblies. Grey segments connecting horizontal lines represent syntenic regions. Orange segments connecting horizonal lines indicate inverted regions. C. 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 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. Significant difference ($p<0.05$) in gene expression between groups is indicated by an asterix (*), as determined by Bonferroni corrected Wilcoxon rank sum test. Complete list of p-values are available in Supplemental Table 3. D. 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. Colours and expression units are the same as in panel B. The complete list of p-values is available in Supplemental Table 4.

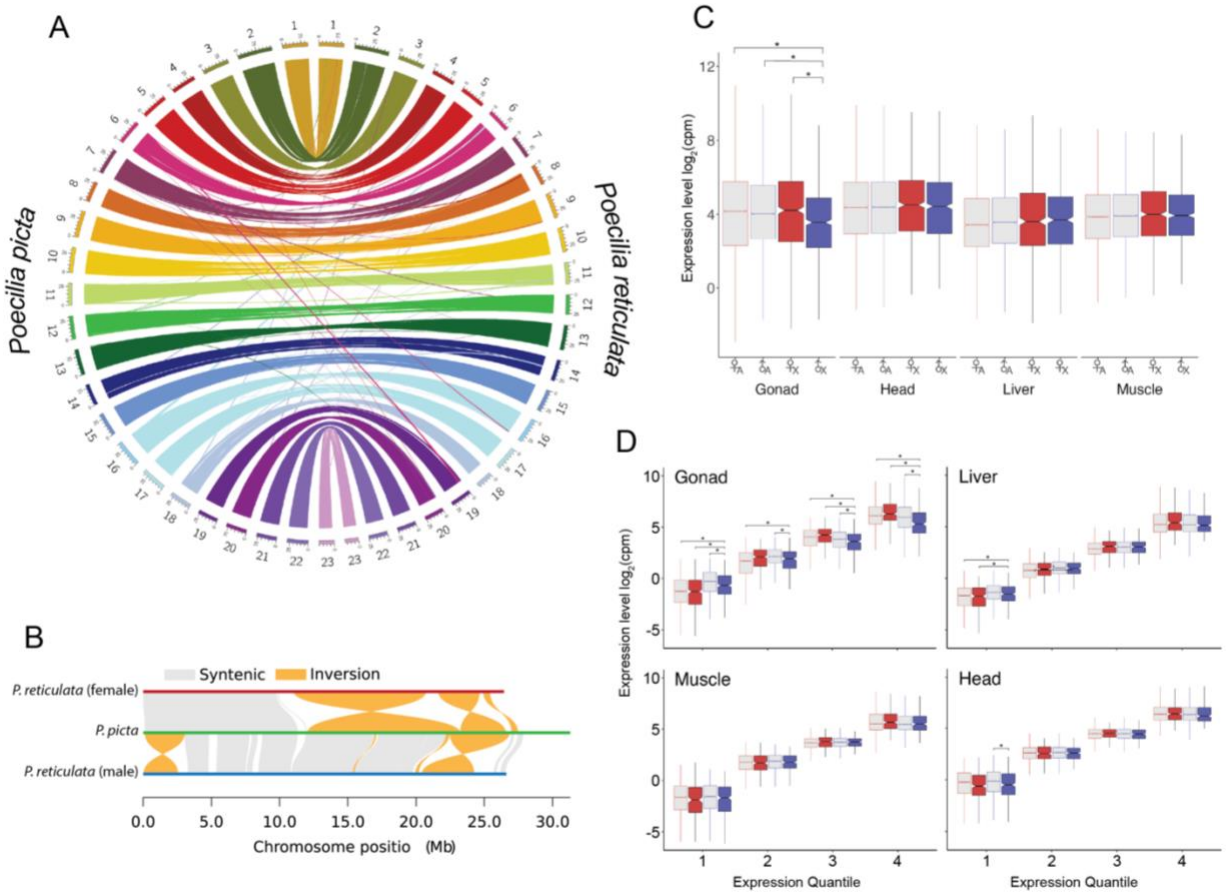
Figure 2. Sex-specific DNA methylation patterns in *P. picta*. Outer ring of circos plot indicates chromosome number and Mb position. Chromosome colours are the same as in Figure 1. Tracks a & b are heatmaps depicting gene (blue) and CpG island (red) densities, each in 100kb bins. Darker colors in the heatmap depict higher counts. Tracks c-d depict histograms of the number of differentially methylated regions (DMRs) in 100kb bins in muscle (orange), liver (green), and gonad (purple) tissues. Histogram bars extending outward represent DMRs that are hypomethylated in males while bars that extend towards the center of the plot represent DMRs that are hypomethylated in females. Chromosome 12 is the X chromosome.

Figure 3. YY1 motif accumulated on the *P. picta* X chromosome from a wave of TE replication and exhibits sex-specific DNA methylation. 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*. Barplots depict the chromosomal distribution of YY1 motifs. Colours represent the different YY1 motif sequences. Nucleotides highlighted in red represent the nucleotide positions that vary between motifs. The phylogeny and divergence times are taken from The Fish Tree of Life,[97] and adapted from[25] and[15]. B. Bar plots depicting the number of YY1 motif containing TEs (gray), YY1 motifs (green), male specific hypomethylated regions in muscle (orange), liver (brown), and gonad (purple) tissue in 100kb bins along the X chromosome. C. DNA methylation levels of genomic regions containing YY1 motifs +/- 2500bps in females (red) and males (blue) on the X (upper panel) and autosomes (bottom panel) presented as a loess regression with 95% confidence intervals in gray. Tissue

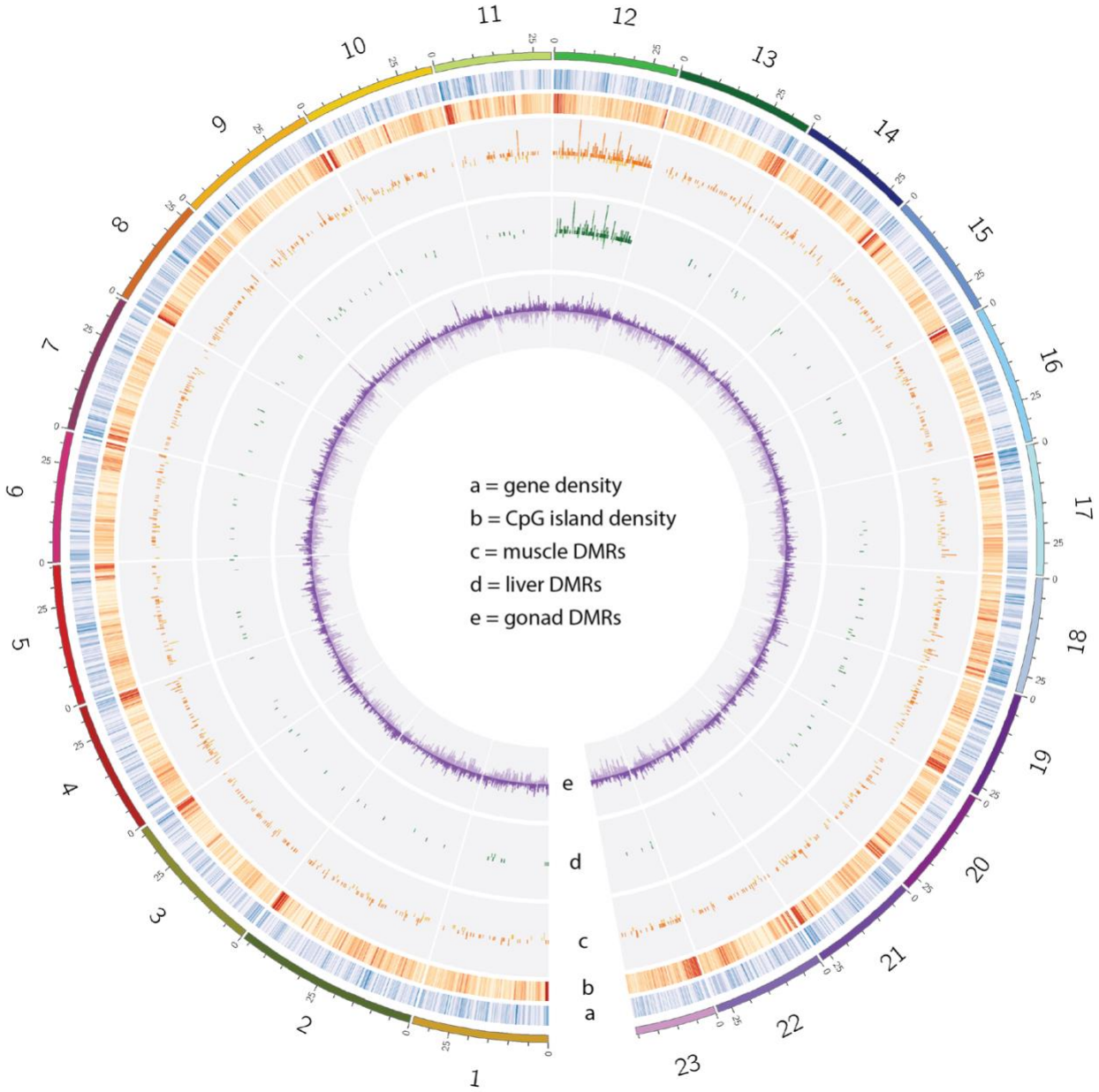
specific DNA methylation indicated by line type; muscle = solid line, liver = dashed line, gonad = dotted line. Zero marks the location of a YY1 motif. D. Circular phylogeny of repetitive sequences containing at least one YY1 motif. Chromosomes are indicated by different colours as in Figures 1 and 2 with the X chromosome in green. Green bars indicate the three clades where sequences are almost exclusively derived from the X chromosome. E. Relationship between the length and number of repetitive elements on each chromosome. Each point represents mean values in 200kb bins for each chromosome (X = green). Length is calculated as the total length of repetitive elements in a 200kb bin. Length represents the sum of the lengths for every repetitive element in a 200kb bin. Confidence intervals are standard error. Count represents the total number of repetitive elements in 1Mb bins. Shaded region is the 95% confidence interval line around the linear regression (solid black line).

Figure 4. **Proposed model for complete X chromosome compensation in *P. picta.***
Hypomethylation of YY1 DNA binding motifs in males promotes male-specific binding of the YY1 transcription factor and the global upregulation of genes on the X chromosome in males. Hypermethylation of YY1 DNA motifs in females prevents YY1 binding and the hyperexpression of X chromosome genes 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 Figure 3.

Figure 1


A
Poecilia picta
Poecilia reticulata
B
Syntenic
Inversion
P. reticulata (female)
P. picta
P. reticulata (male)
0.0
5.0
10.0
15.0
20.0
25.0
30.0
Chromosome positio (Mb)
C
Expression level log2(cpm)
Gonad
Head
Liver
Muscle
D
Gonad
Liver
Muscle
Head
Expression level log2(cpm)
Expression Quantile
Expression Quantile

1088 Figure 2



1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100

Figure 3

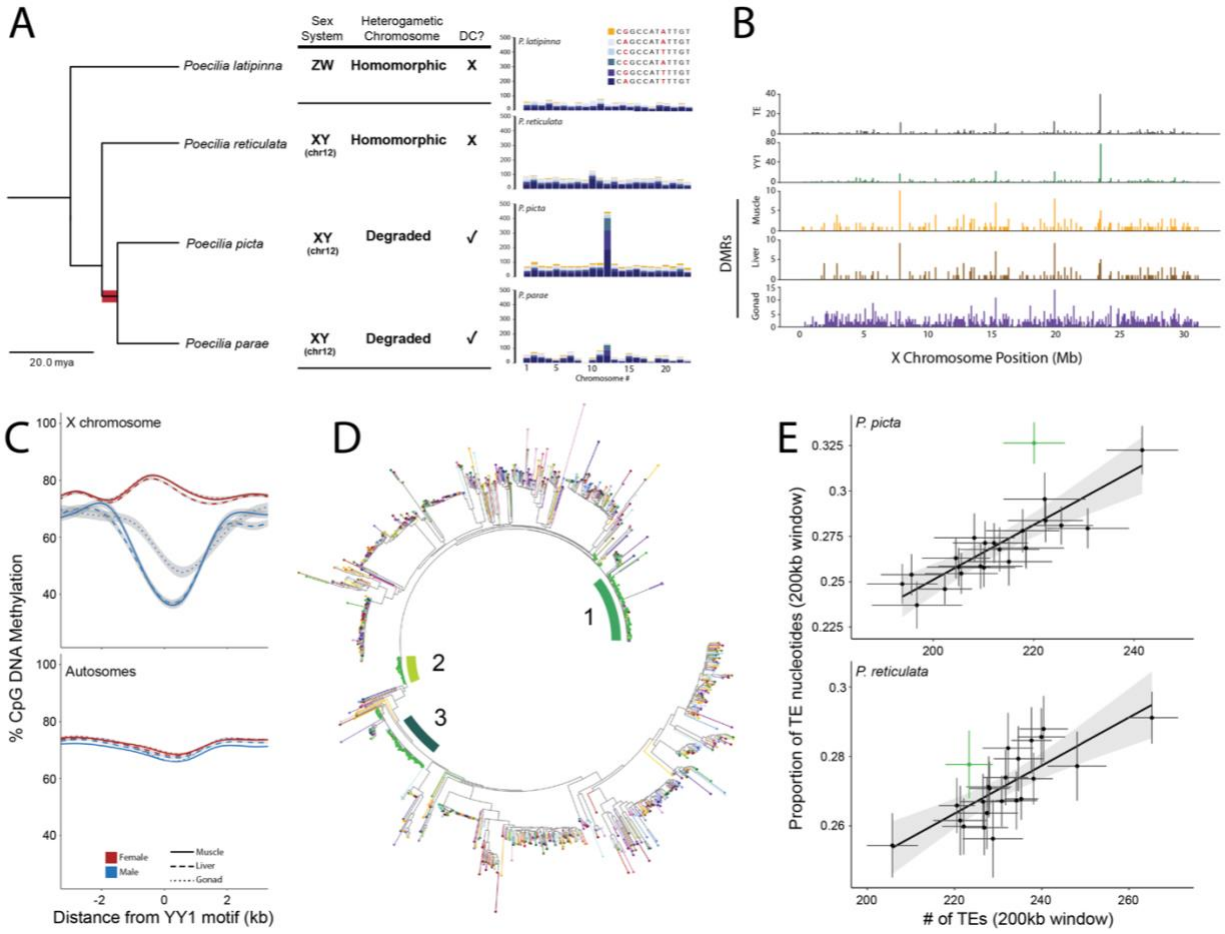
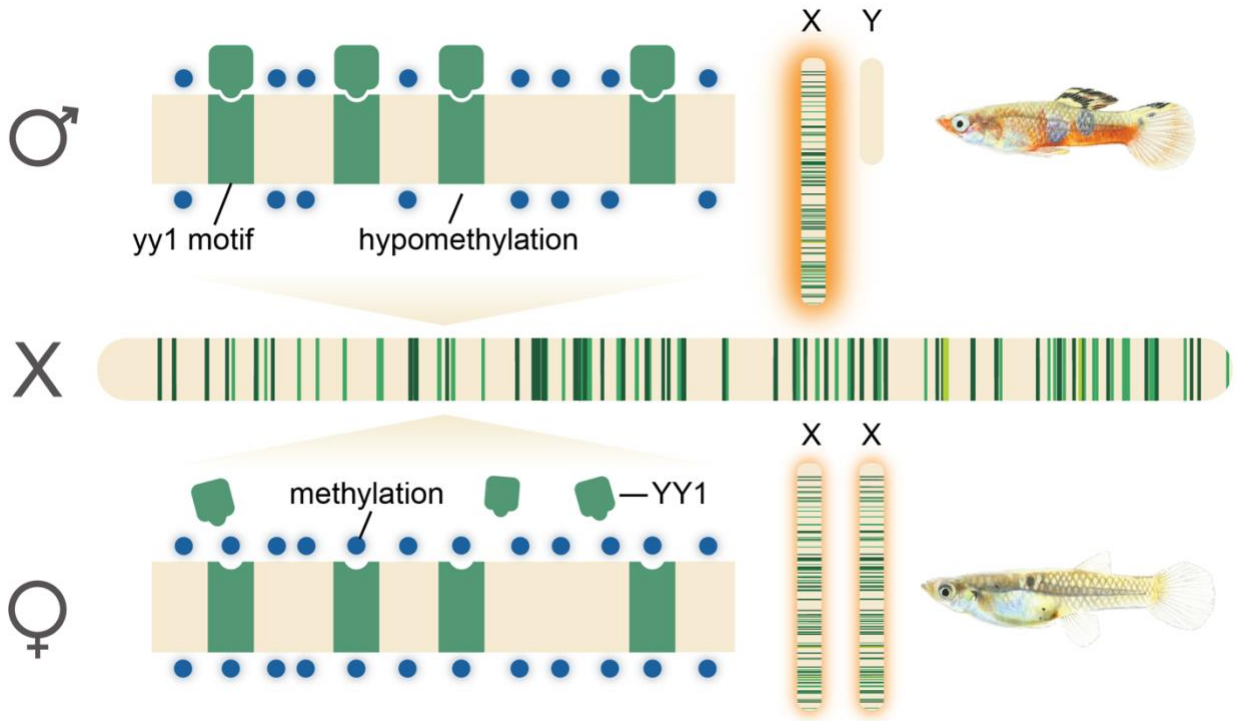
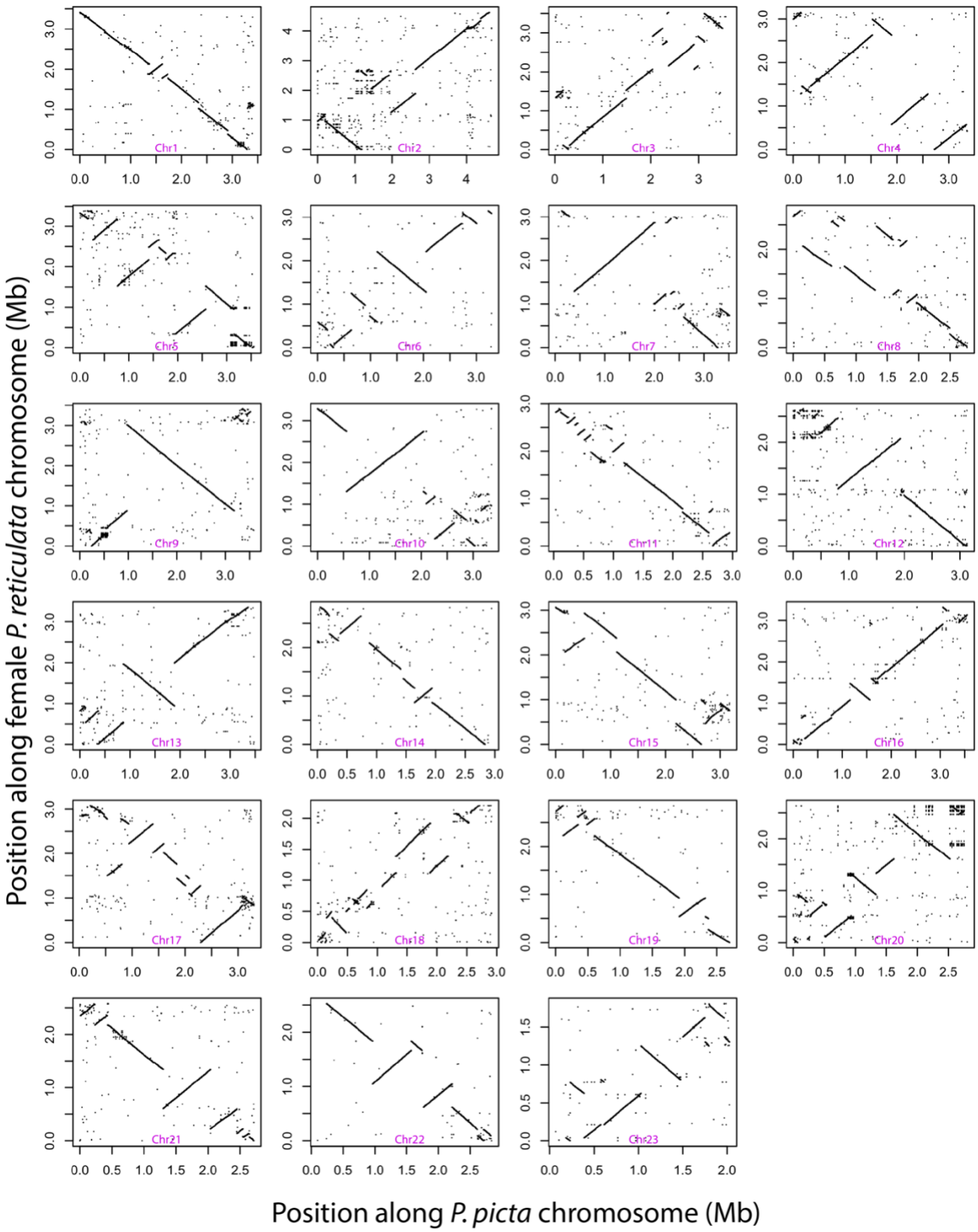
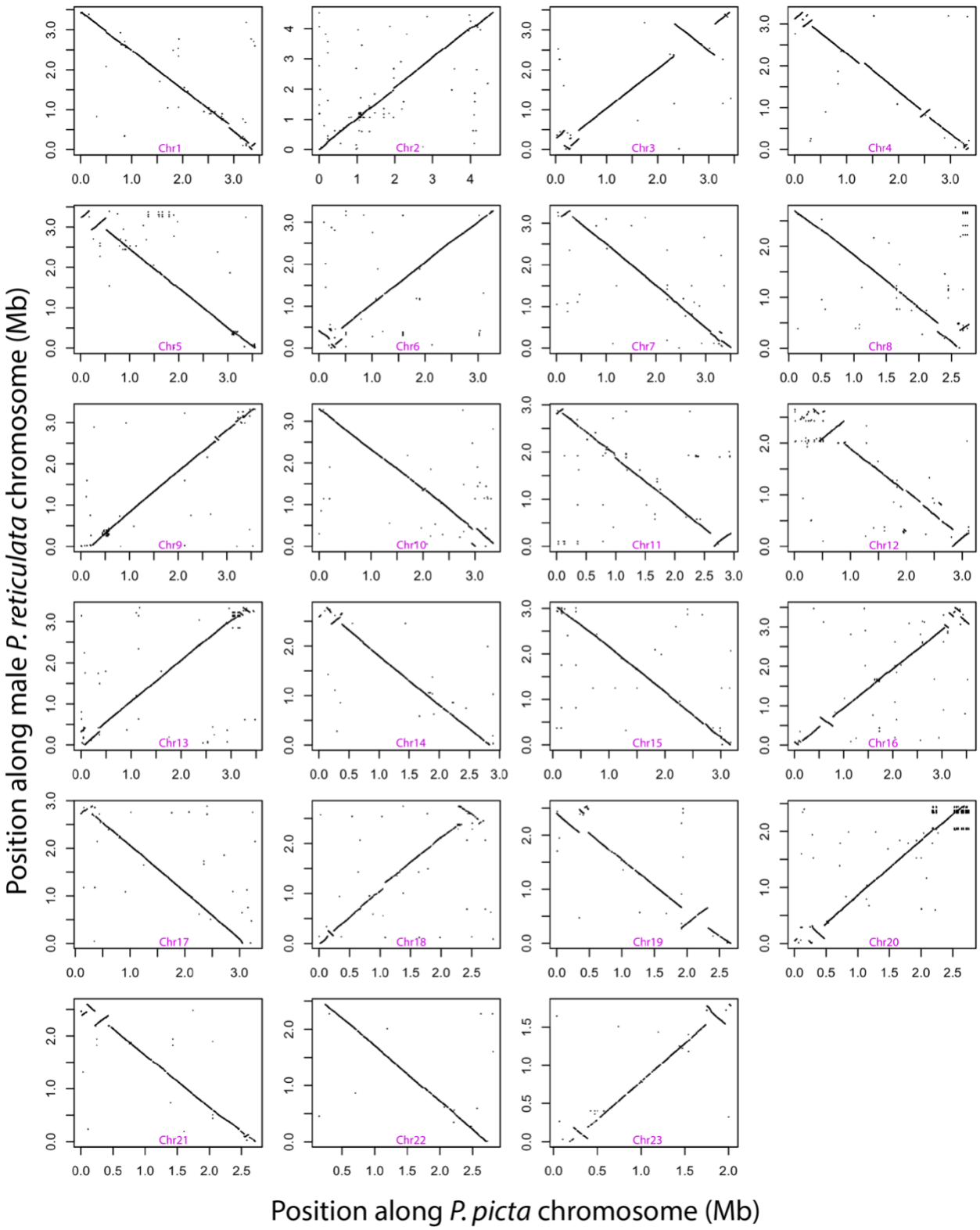

A
Sex System
Heterogametic Chromosome
DC?
Poecilia latipinna
ZW
Homomorphic
X
Poecilia reticulata
XY (chr12)
Homomorphic
X
Poecilia picta
XY (chr12)
Degraded
✓
Poecilia parae
XY (chr12)
Degraded
✓
20.0 mya
P. latipinna
P. reticulata
P. picta
P. parae
Chromosome #
B
TE
YY1
DMRs
Muscle
Liver
Gonad
X Chromosome Position (Mb)
C
X chromosome
Autosomes
% CpG DNA Methylation
Female
Male
Muscle
Liver
Gonad
Distance from YY1 motif (kb)
D
1
2
3
E
P. picta
P. reticulata
Proportion of TE nucleotides (200kb window)
# of TEs (200kb window)

Figure 4

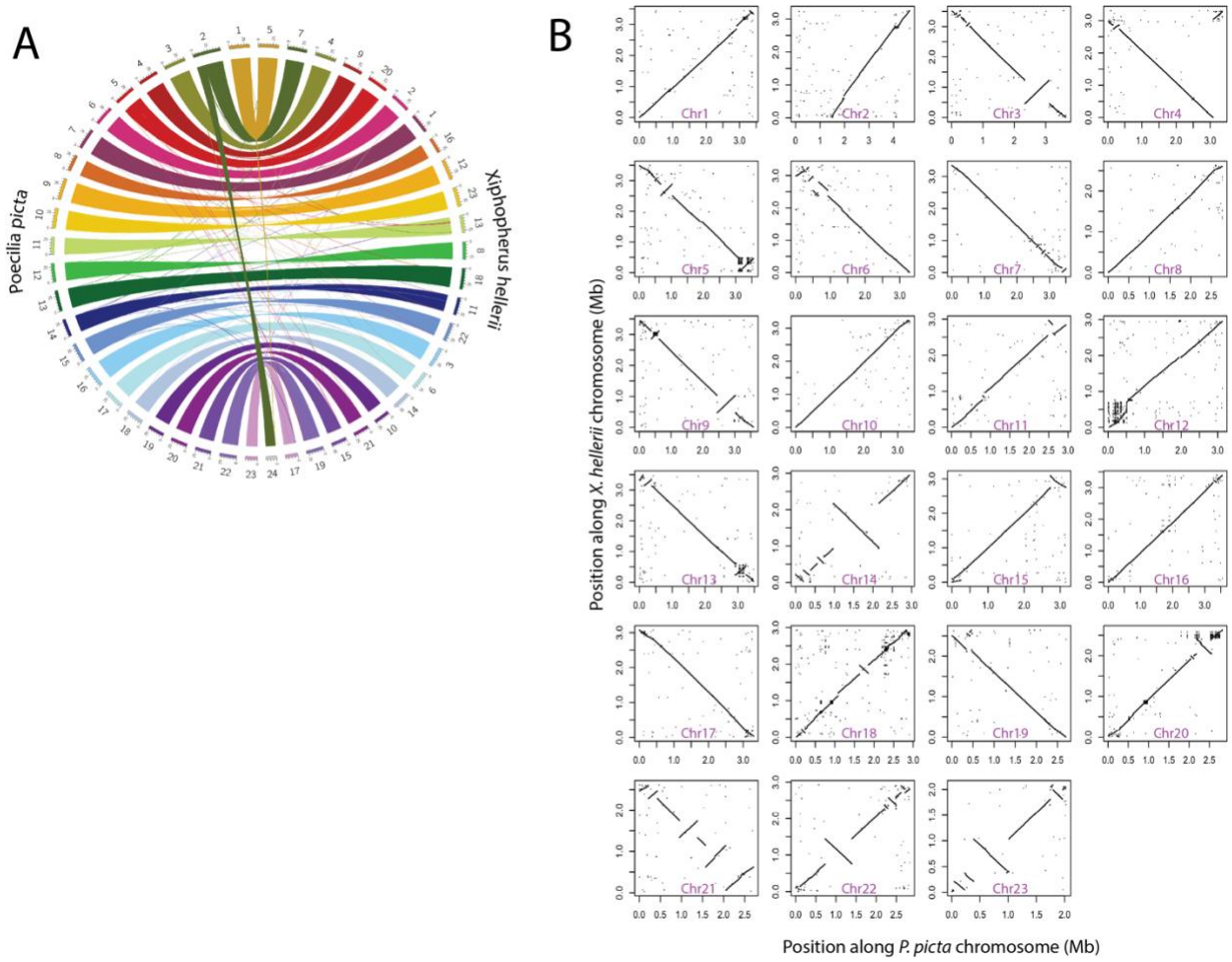




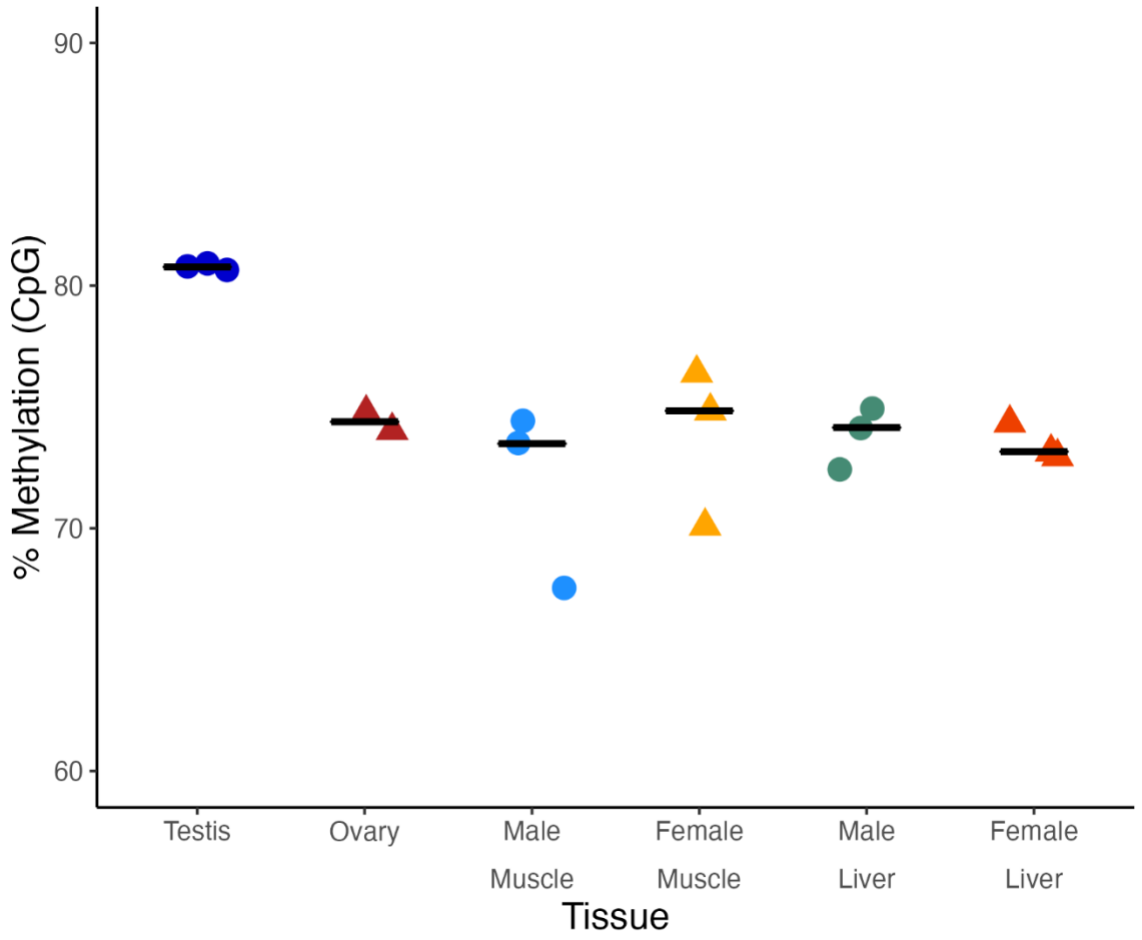
Supplemental Figure 1. Large-scale rearrangements between our *P. picta* and female *P. reticulata* genome assemblies. The X chromosome is Chr12.



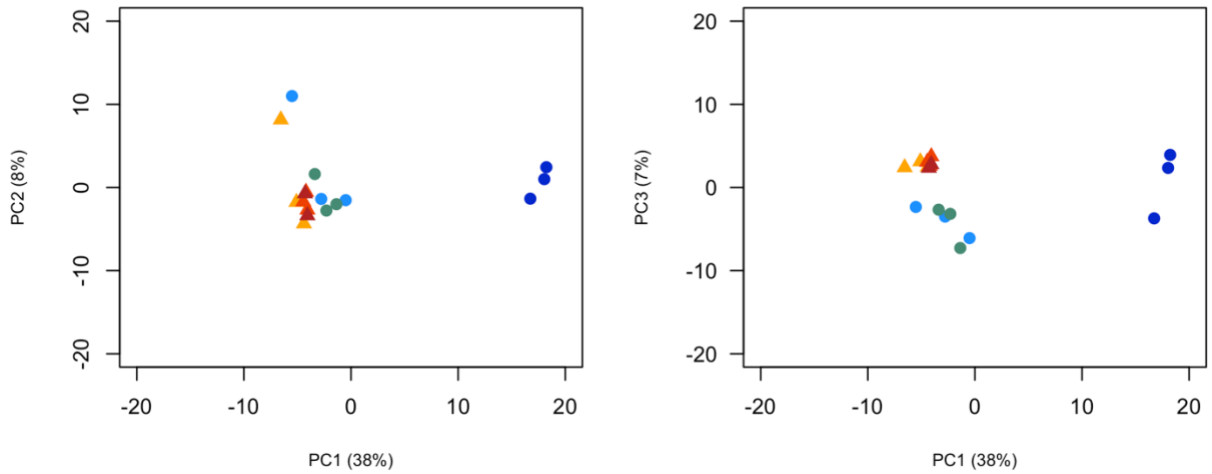
Supplemental Figure 2. Substantial synteny between our *P. picta* and the male *P. reticulata* genome assemblies. The X chromosome is Chr12.



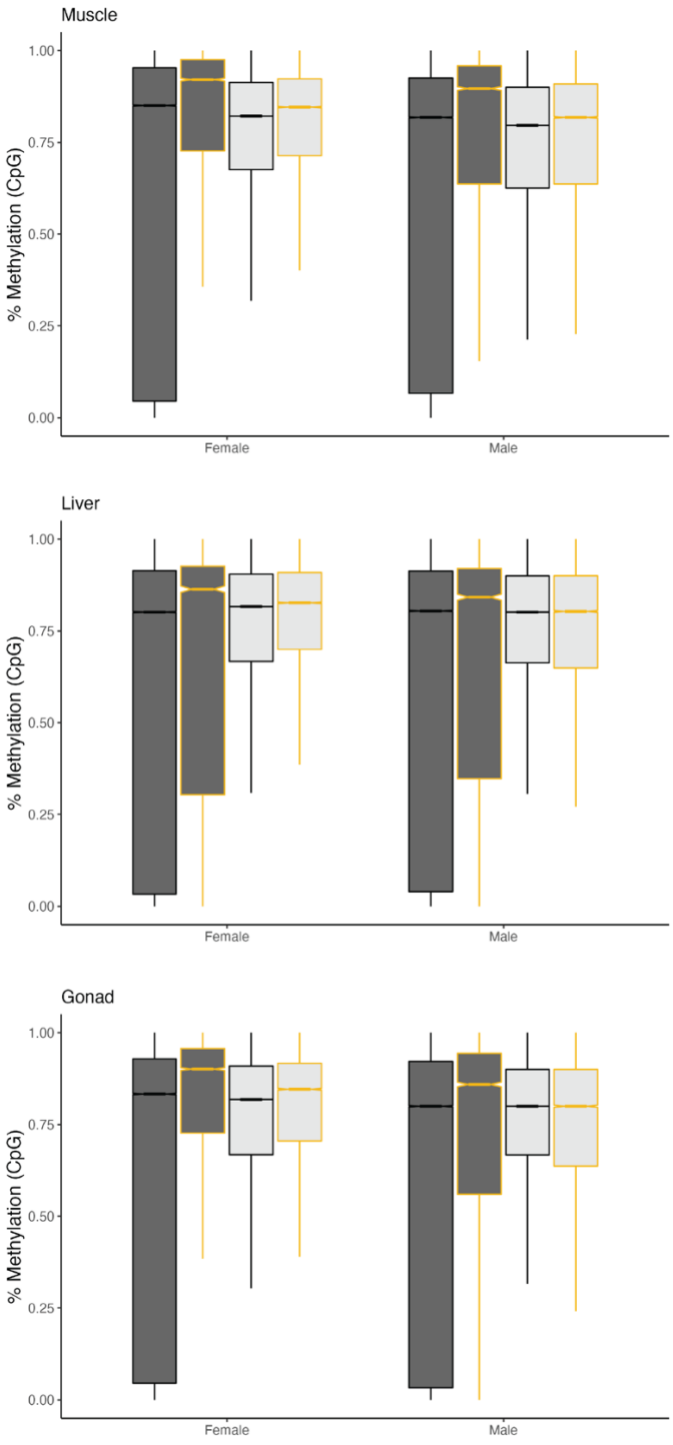
Supplemental Figure3: Sequence comparison between our *P. picta* genome and the *X. hellerii* genome. A) Sequence synteny between the female *P. picta* and the female *P. helleri* genomes presented as a Circos plot. Chromosomes are labeled 1-23 and are differentiated by colour around the outer ring with *P. picta* chromosomes on the left and *P. helleri* chromosomes on the right (numbered 1-24). The colour of the connecting lines indicates the chromosomal origin of the sequence from the *P. picta* genome. B) Synteny scatter plot between *P. picta* and *X. helleri* chromosomes. Each plot is a separate chromosome with the *P. picta* chromosome indicated on each plot. The *P. picta* X chromosome is chr12.



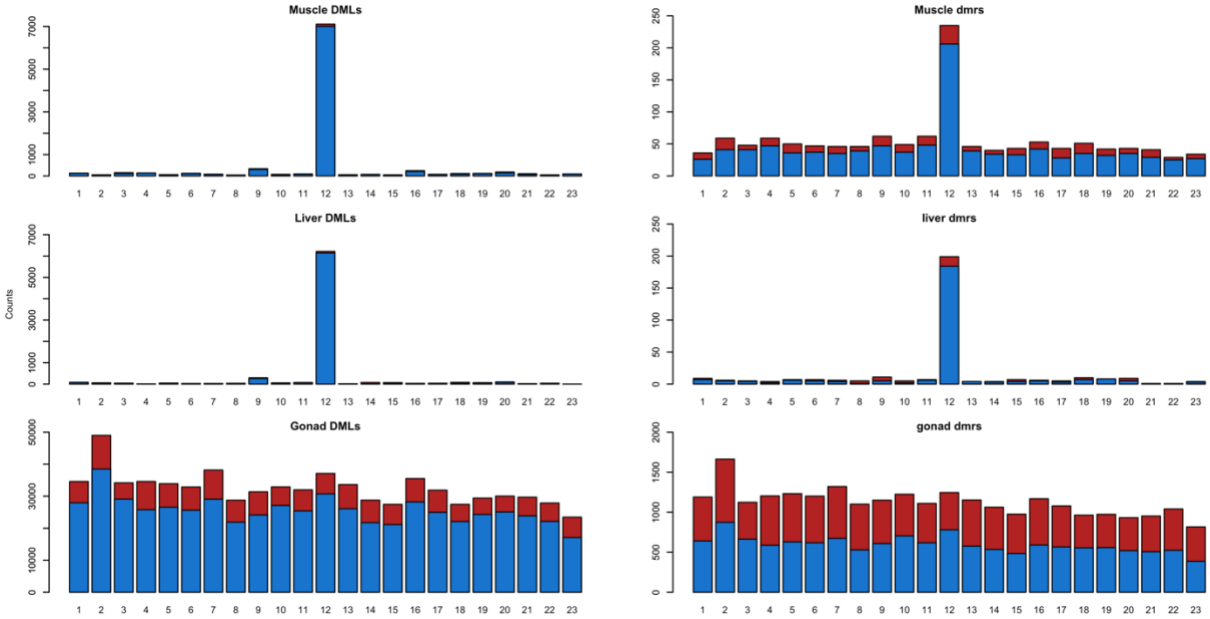
Supplemental Figure 4. Methylation varies by tissue and sex. Whole genome DNA methylation levels for *P. picta* testis (T), ovary (O), male muscle (MM), female muscle (FM), male liver (ML, and female liver (FL) tissue in P. picta. Circles represent individual data points from male tissue. Triangles represent individual datapoints from female tissue. Percent DNA methylation values were calcualted from whole genome bisulfite sequencing data for sites with >10 reads.



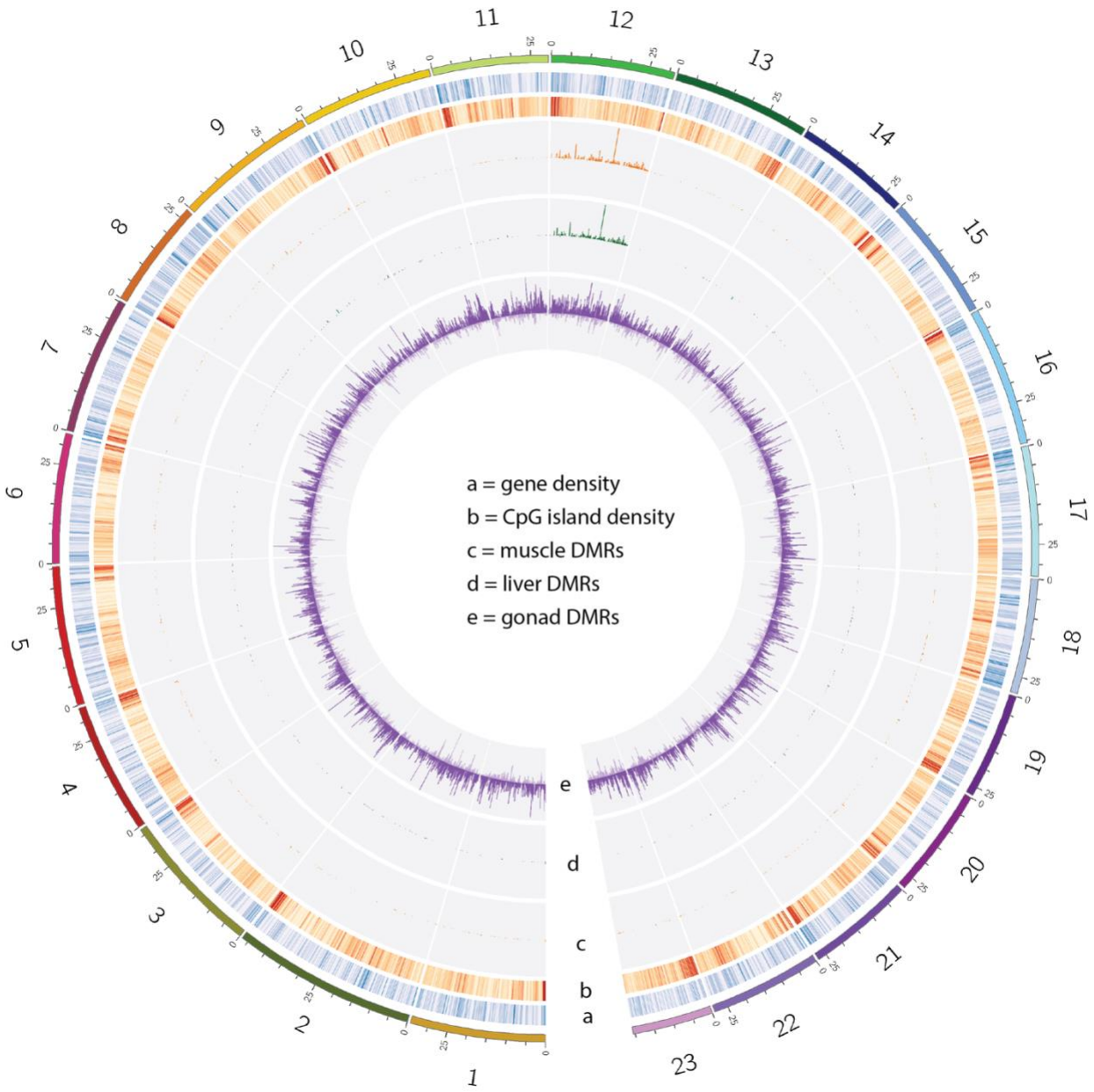
Supplemental Figure 5. Testis is a methylation outlier. Principle componenet analysis of *P. picta* DNA methylation data calculated from whole genome bisulfite sequencing data. Circles represent samples from male tissue. Triangles represent samples from female tissue. Colors are the same as in Supplemental Figure 4.



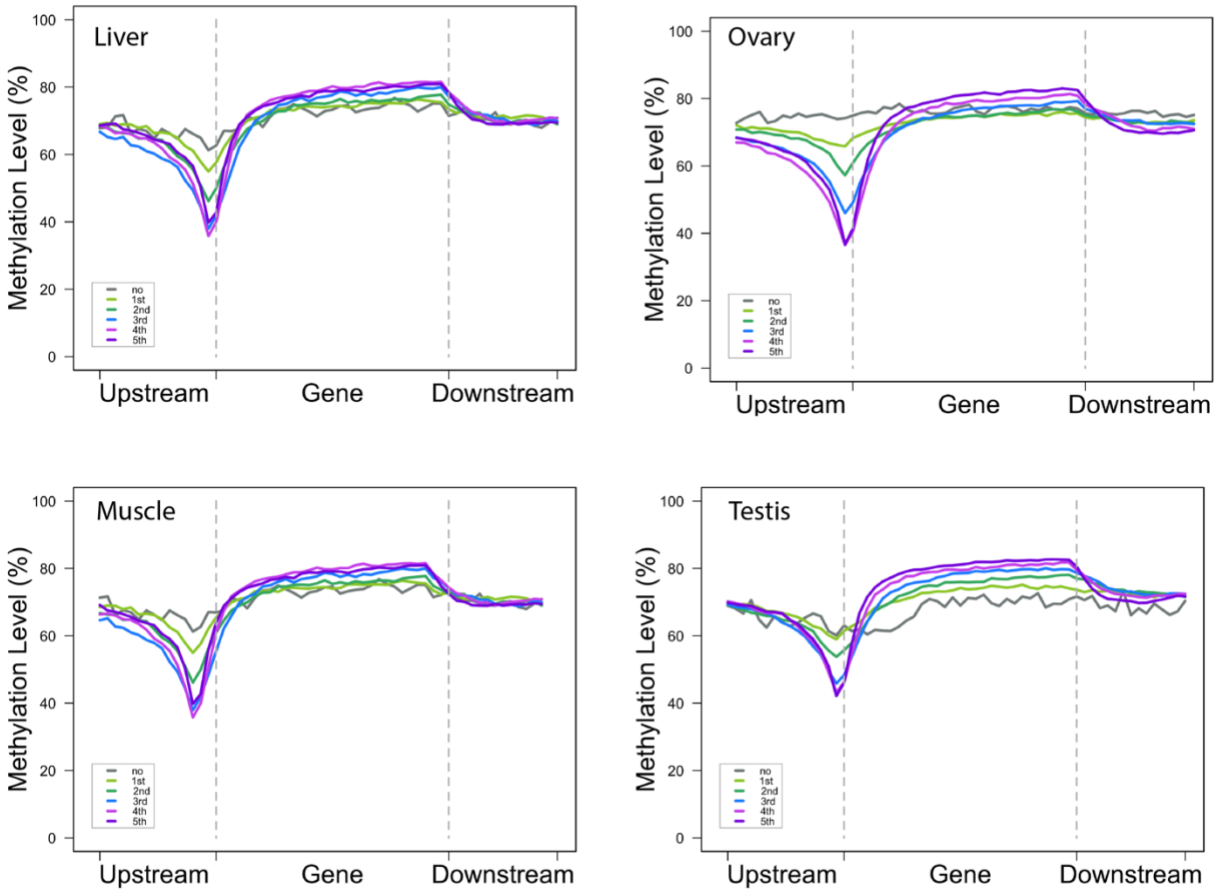
Supplemental Figure 6. CpG islands on the X chromosome are more highly methylated compared to autosomes in both males and females. *P. picta* DNA methylation values of individual CpG located in CpG islands (dark grey) or outside of CpG islands (light grey) on the sex chromosome (orange outline) or autosomes (black outline).



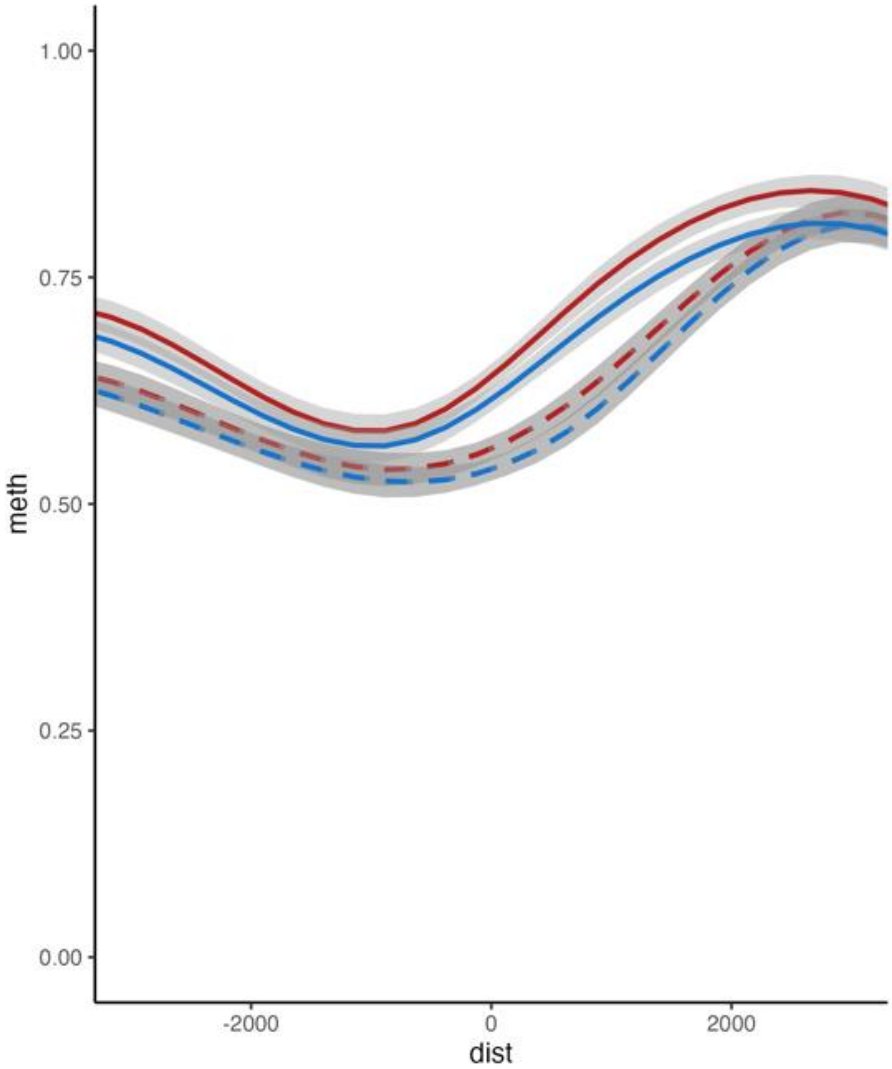
Supplemental Figure 7. The X chromosome is hypomethylated in *P. picta* males. Chromosomal distribution of differntially methylated loci (DMLs) and differentially methylated regions (DMRs) in muscle, liver, and gonad tissue presented as stacked bar plots where the height is the total number of DMLs or DMRs identified on an individual chromosome. Red shaded regions are DMLs or DMRs that are hypomethylated in female tissue Blue shaded regions are DMLS or DMRs that are hypomethylated in male tissue.



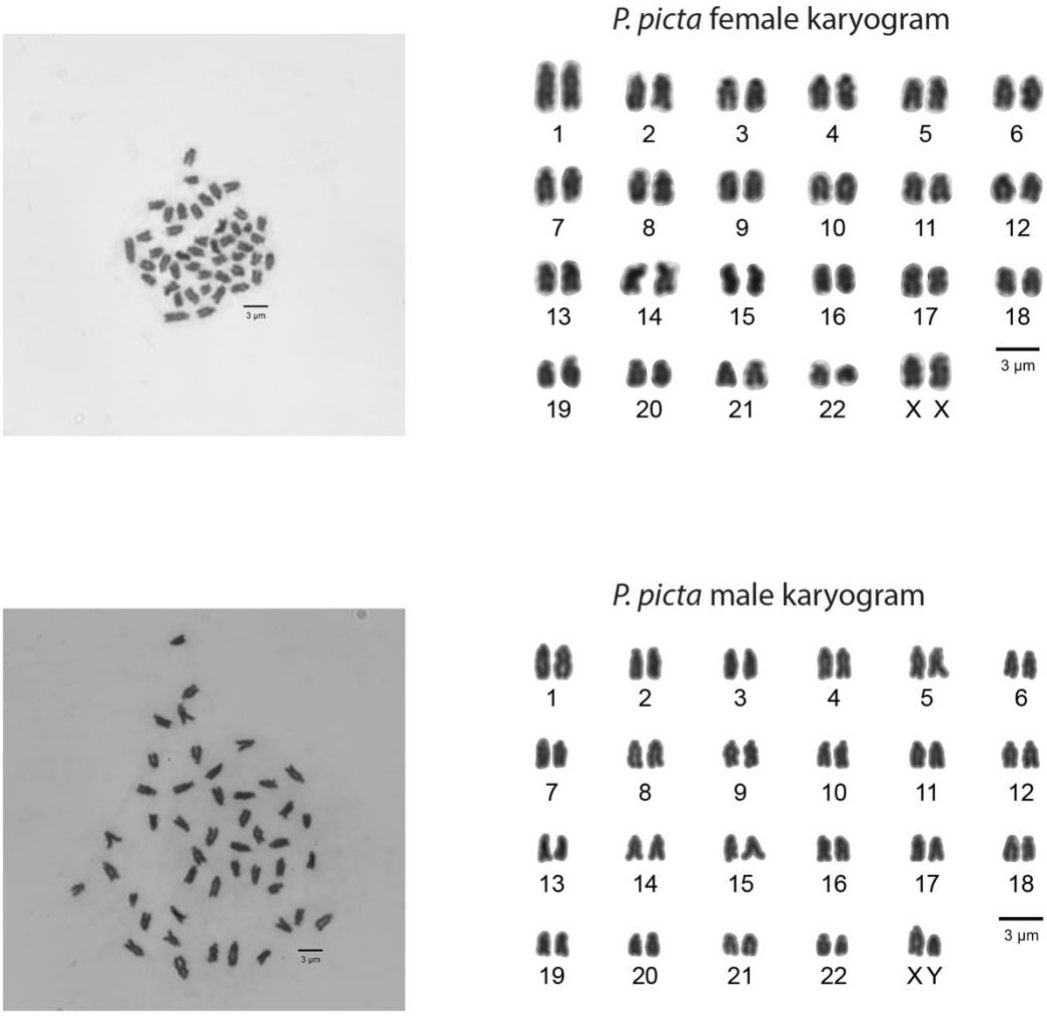
Supplemental Figure 8. Chromosome 12, the X chromosome, is hypomethylated in *P. picta* male somatic tissues. Circos plot depicting sex-specifc DNA methylation patterns in *P. picta*. Outer ring indicates the chromosome number and position in megabases. Chromosome colours are the same as in Figure 1. Tracks a & b are heatmaps depicting gene (blue) and CpG island (Red) densities respectively, in 100kb bins. Darker colors in the heatmap depict higher counts. Tracks c-d depict histograms of the number of differentially methylated CpG loci (DMLs) in 100kb bins in muscle (orange), liver (green), and gonad (purple) tissues respectively. Histogram bars extending outward represent DMLs that are hypomethylated in males while bars that extend towards the center of the plot represent DMLs that are hypomethylated in females.



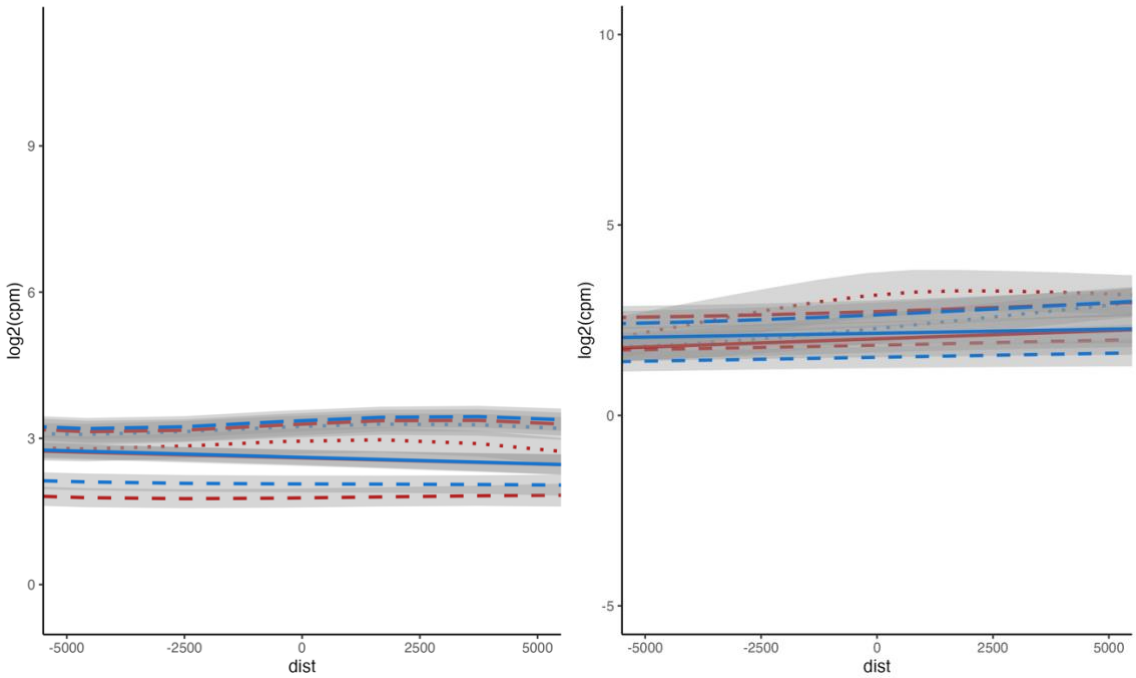
Supplemental Figure 9. Association of DNA methylation and genes ranked by gene expression level in *P. picta*. DNA methylation values for genes with no expression are in gray, genes with low expression are in green, medium expression in blue and pink, and the highest expression in purple.



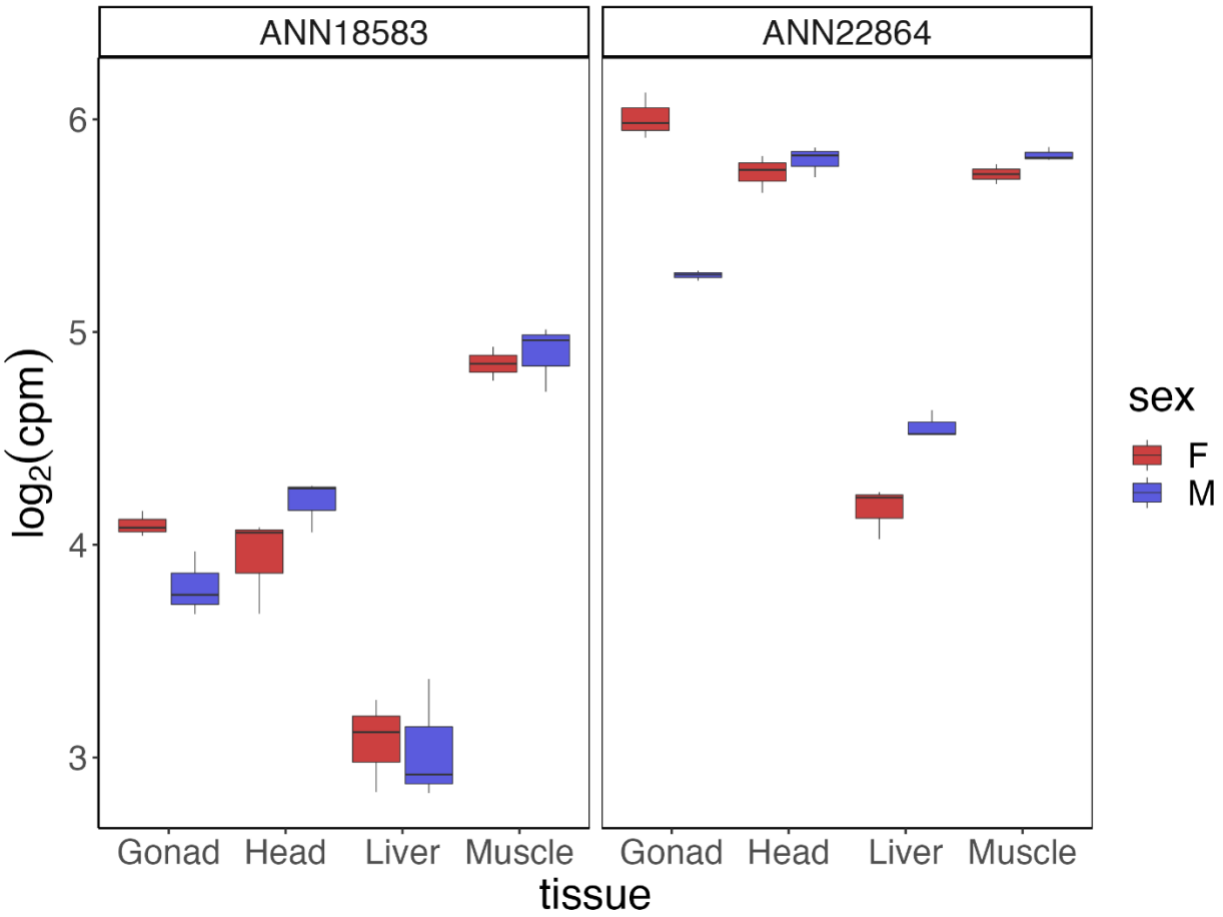
Supplemental Figure 10. CpG islands in gene promoters on the X do not exhibit sex-biased differential methylation. DNA methylation values in the promoter region of genes on the X chromosome that contain CpG islands for female (red) and male (blue) muscle (solid line), liver (dashed line), and gonad (dotted line). Shaded grey region denotes 95% confidence intervals.



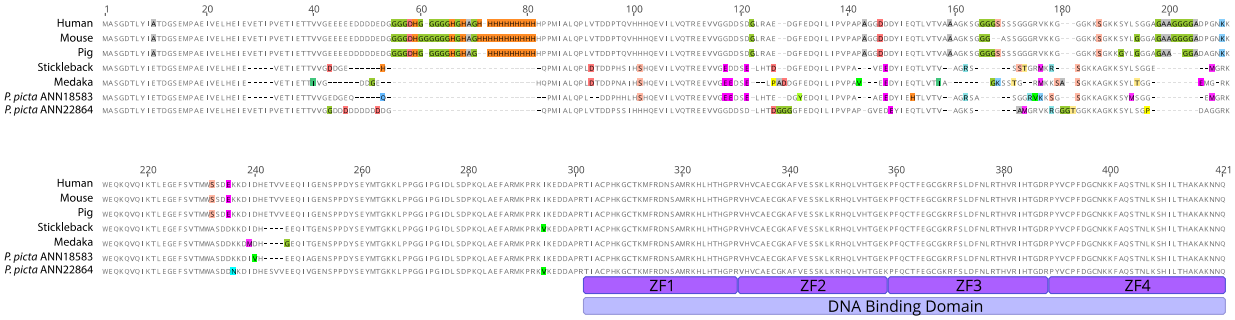
Supplemental Figure 11. Male and female X chromosome appear similar in size. Karyograms from male and female *P. picta* chromosome spreads from our laboratory population originally sourced from Suriname. The Y chromsome appears degenerated compared to the X, while the X chromsomes in males and females are similar in size at this resolution. The left panel shows the chromosome spreads from which the karyograms were sourced.



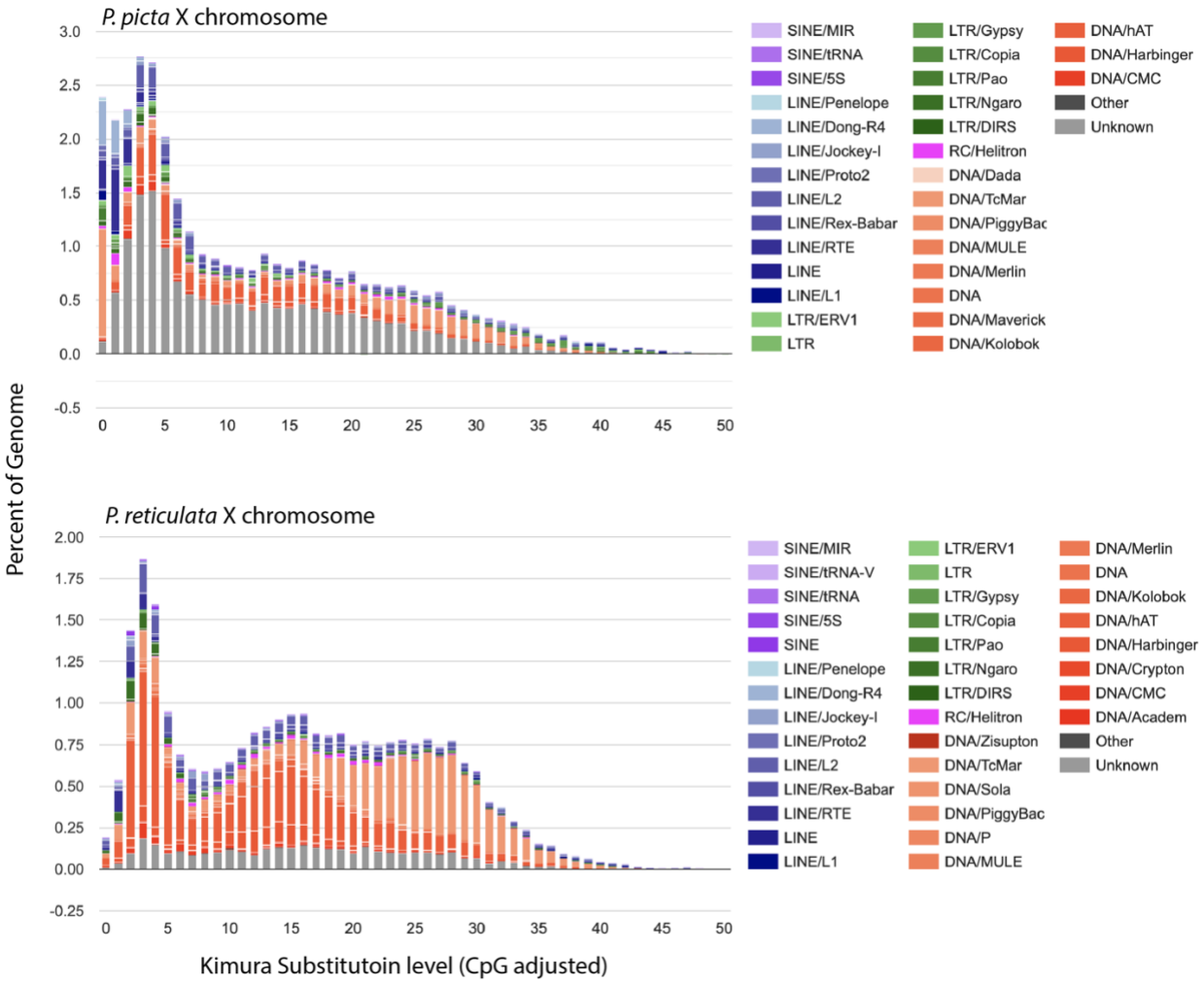
Supplemental Figure 12. Proximity to YY1 motifs is not associated with sex-biased gene expression. Gene expression values for genes 5kb upstream and 5kb downstream of YY1 motifs on the sex chromosome (left panel) or on autosomes (right panel) in female (red) and male (blue) muscle (solid line), liver (dashed line), and gonad tissue (dotted line). Shaded grey areas represent 95% confidence intervals.



Supplemental Figure 13. YY1 orthologs are ubiquitously expressed in head, muscle, liver, and gonad tissues. Gene expression values for two YY1orthologs (ANN18583 and ANN22864) in the *P. picta* genome for female (F, red), and male (M, blue) gonad, head, liver, and muscle tissue. Data is present as log2 counts per million normalized for library size as box and whisker plots. 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.



Supplemental Figure 14 The DNA binding domain of the YY1 protein is conserved between mammals and fishes. Alignment of amino acid sequecnes from human, mouse, pig, stickleback, medaka, and two isoforms from *P. picta*. Coloured amino acid residues depict mismatches between sequences. Purple ZF blocks indicate the four C2H2-type zinc finger domains of the YY1 DNA binding domain.



Supplemental Figure 15. The *P. picta* genome contains a group of recently derrived repetative elements compared to *P. reticulata*. Kimura substitution values used as a proxy for the expansion history of repretitive elements for the *P. picta* (upper panel) and *P. reticulata* (lower panel) X chromosomes. Bargraph depicts the proportion of the X chromosome that contains repetitive elements for a given kimura substitution level. Colors indicate RE annotations. The older the repetitive element expansion the higher the Kimura substitution level indicating more sequence divergence from the consensus.

Supplemental Table 1. *P. picta* genome assembly statistics

HiRise assembly size (bp)	743,972,568
Number of Scaffolds	301
N50	33,053,486
L50	11
N90	27,142,806
L90	21
Largest scaffold	46,707,787
Genome completeness (BUSCO)	97.65 %
Total number of genes	27,764
Total coding region	46,815,666
Repeat %	29.55 %

Supplemental Table 2. Comparison of average expression between the sex chromosome and autosomes across tissues and sex.

Tissue	X	Y	p-value*
Gonad	Male X genes	Male Autosomal genes	4.915e-12
		Female X genes	6.443e-07
		Female Autosomal genes	1.036e-07
Muscle	Male X genes	Male Autosomal genes	0.6724
		Female X genes	0.7923
		Female Autosomal genes	0.2407
Liver	Male X genes	Male Autosomal genes	0.9855
		Female X genes	0.6603
		Female Autosomal genes	0.1029
Head	Male X genes	Male Autosomal genes	0.94
		Female X genes	0.3903
		Female Autosomal genes	0.8463

*p-values calculated using Wilcoxon rank sum test between data in column X and data in column Y

Supplemental Table 3. Wilcoxon rank sum test statistics for dosage compensation as a function of expression level.

Tissue	Quantile	X	Y	p-value*
Gonad	Quantile 1	Male X genes	Male Autosomal genes	3.503e-04
			Female X genes	1.919e-07
			Female Autosomal genes	2.9e-15
	Quantile 2	Male X genes	Male Autosomal genes	0.002124088
			Female X genes	1
			Female Autosomal genes	0.02179697
	Quantile 3	Male X genes	Male Autosomal genes	0.01034029
			Female X genes	4.166e-08
			Female Autosomal genes	1.44e-05
	Quantile 4	Male X genes	Male Autosomal genes	3.305e-07
			Female X genes	1.892e-11
			Female Autosomal genes	5.725e-10
Muscle	Quantile 1	Male X genes	Male Autosomal genes	0.8039929
			Female X genes	0.3912101
			Female Autosomal genes	1
	Quantile 2	Male X genes	Male Autosomal genes	1

			Female X genes	1
			Female Autosomal genes	1
	Quantile 3	Male X genes	Male Autosomal genes	1
			Female X genes	1
			Female Autosomal genes	0.6628782
	Quantile 4	Male X genes	Male Autosomal genes	1
			Female X genes	0.8554198
			Female Autosomal genes	1
Liver	Quantile 1	Male X genes	Male Autosomal genes	0.2517831
			Female X genes	0.0006019065
			Female Autosomal genes	0.00002378624
	Quantile 2	Male X genes	Male Autosomal genes	1
			Female X genes	1
			Female Autosomal genes	0.02162681
	Quantile 3	Male X genes	Male Autosomal genes	1
			Female X genes	1
			Female Autosomal genes	0.1268024
	Quantile 4	Male X genes	Male Autosomal genes	1

			Female X genes	0.6283786
			Female Autosomal genes	1
Head	Quantile 1	Male X genes	Male Autosomal genes	0.008493841
			Female X genes	1
			Female Autosomal genes	0.9011067
	Quantile 2	Male X genes	Male Autosomal genes	1
			Female X genes	1
			Female Autosomal genes	1
	Quantile 3	Male X genes	Male Autosomal genes	1
			Female X genes	1
			Female Autosomal genes	1
	Quantile 4	Male X genes	Male Autosomal genes	0.5104976
			Female X genes	0.7668176
			Female Autosomal genes	0.2904713

* Bonferroni corrected p-value results from Wilcoxon rank sum test between data in column X and data in column Y

Supplemental Table 4. RNA-Seq data for *P. picta.*

Tissue	Sex	Biological Replicate	# reads	# reads mapped as pairs	% reads mapped
Muscle	Male	1	144636107	118061726	88.28%
		2	169476423	138122808	89.86%
		3	194995685	158505438	90.33%
	Female	1	148583536	120583282	91.11%
		2	165442661	130740620	89.83%
Liver	Male	1	42653690	42602048	99.88%
		2	46358957	46311114	99.90%
		3	42446289	42401986	99.90%
	Female	1	41777391	41741864	99.91%
		2	43934472	43895938	99.91%
		3	42383097	42341376	99.90%
Head	Male	1	51631056	51544142	99.83%
		2	46551675	46479572	99.85%
		3	38206399	38148536	99.85%
	Female	1	54132041	54053660	99.86%
		2	60580208	60497298	99.86%
		3	46450145	46377666	99.84%
Testis	Male	1	45585537	45535708	99.89%
		2	59703504	59643374	99.90%
		3	51146625	51079088	99.87%
Ovary	Female	1	53248762	53146800	99.81%
		2	51869747	51785328	99.84%
		3	57592595	57497696	99.84%

Supplemental Table 5. WGBS data for *P. picta.*

Tissue	Sex	Biological Replicate	# reads	# reads mapped as pairs	% reads mapped
Muscle	Male	1	184771950	125246074	89.75%
		2	160032514	112189848	88.53%
		3	160032514	112189848	88.53%
	Female	1	173524325	113572030	89.81%
		2	143912326	110264156	91.91%
		3	166238590	119300796	91.08%
Liver	Male	1	151693943	101210542	87.24%
		2	175369447	103025500	80.21%
		3	173169280	115526270	87.01%
	Female	1	177598077	113951140	91.27%
		2	158676444	110121160	90.81%
		3	133363848	92716046	90.58%
Testis	Male	1	165429956	87663290	71.92%
		2	143549255	76288622	72.66%
		3	130382202	68272628	76.52%
Ovary	Female	1	172253818	121325294	90.46%
		2	156043172	113553024	91.51%