Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation

Alison E. Wright[1], Iulia Darolti[1], Natasha I. Bloch[1], Vicencio Oostra[1], Ben Sandkam[2], Severine D. Buechel[3], Niclas Kolm[3], Felix Breden[2], Beatriz Vicoso[4] & Judith E. Mank[1]

Sex chromosomes evolve once recombination is halted between a homologous pair of chromosomes. The dominant model of sex chromosome evolution posits that recombination is suppressed between emerging X and Y chromosomes in order to resolve sexual conflict. Here we test this model using whole genome and transcriptome resequencing data in the guppy, a model for sexual selection with many Y-linked colour traits. We show that although the nascent Y chromosome encompasses nearly half of the linkage group, there has been no perceptible degradation of Y chromosome gene content or activity. Using replicate wild populations with differing levels of sexually antagonistic selection for colour, we also show that sexual selection leads to greater expansion of the non-recombining region and increased Y chromosome divergence. These results provide empirical support for longstanding models of sex chromosome catalysis, and suggest an important role for sexual selection and sexual conflict in genome evolution.

[1] Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. [2] Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6. [3] Department of Zoology, Stockholm University, Svante Arrheniusväg 18 B, Stockholm 106 91, Sweden. [4] Institute of Science and Technology, Am Campus 1A, Klosterneuburg 3400, Austria. Correspondence and requests for materials should be addressed to A.E.W. (email: alison.e.wright@ucl.ac.uk).

Sex chromosomes are typically thought to evolve as recombination is halted between a homologous pair of chromosomes in one sex. Although we have a detailed understanding of the evolutionary consequences of the loss of recombination for sex chromosome evolution[1,2], we still do not understand the evolutionary forces acting to halt recombination in the first place. The dominant theoretical model for the early stages of sex chromosome evolution[3–5] predicts that recombination will be selected against in the region between a sex determining gene and a nearby locus with alleles of sex-specific effect. This theory, though prevalent, remains largely untested empirically, as most research has focused on older, highly divergent sex chromosome systems[6,7], for which it is difficult to extrapolate the earliest stages and causes of divergence.

The sex chromosomes of the guppy (Poecilia reticulata) have been of interest for more than a century, following early reports that many sexually selected colour traits are passed through the patriline on the Y chromosome[8,9]. These observations were central to the development of theories regarding the role of sexual conflict in recombination suppression and sex chromosome divergence[3–5]. Colour is sexually antagonistic in guppies, as brightly coloured males are more attractive to females and more visible to predators, but brightly coloured females gain no fitness advantage and only suffer increased predation[10–12]. Therefore, in this system, current models of sex chromosome evolution predict that recombination would be selected against between the sex determining locus and linked loci involved in colouration. This process would shrink the pseudoautosomal region in favour of expanding X- and Y-specific regions, creating a male supergene on the Y chromosome containing multiple colouration loci and thereby resolving sexually antagonistic selection.

Even though the guppy sex chromosomes are a classic model for the study of sexual conflict and sex chromosome divergence, little is actually known about the pattern of divergence between the X and Y chromosomes. Recent linkage maps identified male recombination events restricted to the middle of chromosome 12 (ref. 13), suggesting that the other half of the chromosome is functionally X- or Y-linked. Immunostaining of recombination nodules[14] was broadly concordant with recombination mapping, again suggesting that the X chromosome is split roughly in equal parts between X-specific and pseudoautosomal.

Recombination shows substantial local variation between males and females throughout the genomes of many organisms[7,15] and identifying areas of restricted male recombination does not distinguish the sex chromosome from other areas where males simply do not recombine. However, the Y is morphologically distinguishable from the X chromosome[16], and comparative genome hybridization of lab populations[17] suggest that roughly half of the Y chromosome is male-specific. Because many vertebrate sex chromosomes show progressive spread of the non-recombining region[18–21], the large size of the guppy non-recombining region and male-specific regions suggest substantial divergence between the X and Y.

Recombination suppression between the X and Y chromosomes results in complete linkage of the male-specific region of the Y. The loss of recombination in this region typically limits the role of adaptive evolution and leads to strong background selection and linkage effects, causing loss of functional polymorphism in coding sequence over time[1]. Roughly half of male colouration patterns are thought to be Y-linked[8], and the remarkable diversity of male colour combinations implies an improbably large number of Y haplotypes maintained within populations for a sex chromosome system of at least intermediate age. Additionally, if recombination suppression really is driven by sexually antagonistic alleles[3–5], then we might expect recent

but rapid spread of recombination suppression shortly after the emergence of sexual preferences for colour. Although sexually selected traits exist in many Poeciliids, the vivid male colouration in P. reticulata is only shared by a few very close relatives[22,23], therefore, the expansions of the male-limited Y chromosome to engulf colouration loci might have occurred very recently.

Moreover, the degree of male colouration, and, therefore, the degree of sexual conflict over colour, varies substantially based on predation pressures. Across watersheds, downstream populations are typically associated with higher predation and males are far less colourful than upstream populations[24–26]. Importantly, the proportion of colour patterns thought to be Y-linked varies between upstream and downstream populations[27]. The unusual gene content of the guppy sex chromosomes, therefore, makes it a uniquely powerful system for testing the role of sexual conflict and sexual selection in sex chromosome divergence.

In order to determine the degree of divergence between the X and Y chromosome in this species, we resequenced male and female genomes and transcriptomes of both laboratory and wild individuals. We find that the X and Y show sequence differentiation over nearly one half of the length of the chromosome, however, the divergence between the X and Y chromosome is remarkably subtle, indicating very low levels of divergence and likely recent origin of the sex chromosome. The large region of divergence is in contrast to reports of other nascent sex chromosome systems where the diverged region is highly restricted[28–30]. Despite this young age, we detect evidence of Faster-X evolution in this region. Most importantly, we find convergent patterns of greater sex chromosome divergence in upstream populations, which experience substantially elevated sexual selection and sexual conflict, compared with downstream populations. Our results suggest that recombination suppression between the X and Y spread quickly in the recent history of this sex chromosome system, possibly driven by the presence of sexually antagonistic alleles related to sexual selection.

Results

The structure of the guppy sex chromosomes. We first assembled the female genome using SOAPdenovo2, based on 480 million paired end reads from an outbred laboratory population. The assembly yielded 96,611 scaffolds, with an N50 of 11.3 Kb and total length of 634.8 Mb, after a minimum length threshold of 1 Kb (Supplementary Tables 1–3). Guppy genes from the reference genome (Guppy_female_1.0 + MT) were mapped to scaffolds in order to identify chromosomal positions, resulting in a final assembly of 19,206 ordered scaffolds oriented along the guppy chromosomes, with an N50 of 17.4 Kb and total length of 219.5 Mb (Supplementary Tables 3 and 4).

We then mapped male and female DNA-seq reads to our ordered scaffolds in order to identify regions of coverage difference between the sexes. Regions with longstanding recombination suppression in males will show reduced mapping efficiency against the female genome assembly, as diverged sequence from the Y will no longer map to the X chromosome[19,20,31]. Even with strict mapping thresholds (see methods) we could identify no large region of the genome with reduced coverage in males, which we would expect if a large portion of the Y was significantly diverged or degraded (Supplementary Fig. 1), and the overall distribution of coverage is largely symmetrical (Supplementary Fig. 2A). However, previous linkage maps have identified chromosome 12, which contains the sex determining gene, as the sex chromosome[13], and this chromosome shows a slight shift in the distribution and has a significantly greater
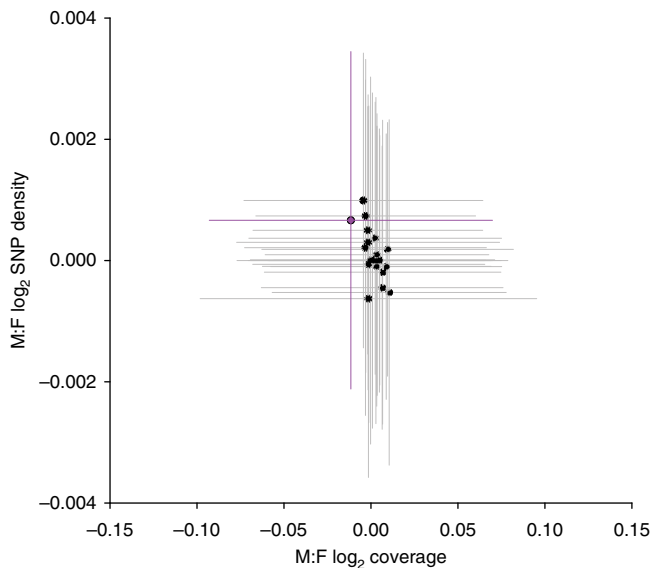
Figure 1 | Distribution of sex differences in coverage and SNP density for all chromosomes. The X chromosome is in purple. Horizontal and vertical lines denote interquartile ranges.

proportion of scaffolds with female-biased coverage than autosomes (Wilcoxon rank sum test $P < 0.001$, Supplementary Fig. 2B, Supplementary Table 5). This suggests that recombination suppression between the X and Y chromosomes has led to very slight divergence between them.

If the Y has diverged, but not yet degraded significantly, we would expect to observe Y-specific single nucleotide polymorphisms (SNPs) in regions that retain substantial sequence similarity to the X, resulting in higher average male heterozygosity for the sex chromosomes[32,33]. When assessing all regions of the genome, we observe a shoulder of elevated SNP density in males (Supplementary Fig. 2C), due to significantly greater SNP density in males for the sex chromosomes compared with autosomal genes (Wilcoxon rank sum test $P < 0.001$, Supplementary Fig. 2D, Supplementary Table 5). When sex differences in coverage and SNP density are plotted together, the sex chromosome is a clear outlier to the other chromosomes (Fig. 1), confirming low but significant levels of divergence.

In order to determine the relative divergence between X and Y chromosomes, we plotted coverage and SNP density differences between males and female on our scaffolds against physical position on the guppy genome assembly. We detected significantly reduced male coverage outside the autosomal 95% confidence interval from 22–25 Mb (Fig. 2). This region shows the largest degree of X-Y sequence divergence and likely corresponds to the oldest region of the sex chromosome (Stratum I). In contrast, between 15–25 Mb, we detect significant elevation of male SNP density but no reduction in male coverage, indicative of lower levels of X-Y divergence and suggesting that nearly half of the sex chromosome has stopped recombining in males in the very recent past (Stratum II) (Fig. 3a, Supplementary Fig. 3).

Our coverage and SNP analysis suggest that although male-specific SNPs have accumulated, the Y chromosome has not degenerated significantly. Because loss of gene activity often quickly follows loss of recombination on the Y chromosome[1], for each gene we plotted male and female expression level (RPKM) across the X chromosome. Our results show that the non-recombining region exhibits low levels of sexualization of

gene content, with regions where the majority of genes exhibit female- or male-biased expression. However, there is no region of detectable loss of male gene activity, as would be expected with extensive Y chromosome decay (Fig. 3b, Supplementary Fig. 4). In contrast, in the region of the sex chromosome with the greatest coverage difference between males and females (Stratum I, 22–25 Mb), likely the area of greatest Y chromosome divergence, there is a slight excess of male-biased genes, indicating that this region of the Y chromosome has also not suffered any significant loss of gene activity. We tested for enrichment of GO terms for genes expressed in the X-Y diverged region (Strata I and II, 15–25 Mb) relative to the rest of the genome. However, there were no GO terms with an enrichment $P < 0.001$.

X chromosomes are predicted in many circumstances to show elevated rates of evolution[34], and signatures of Faster-X evolution have been detected in old, heteromorphic sex chromosomes[35–37]. However, it is unclear whether a detectable signal of Fast-X would be expected in the early stages of sex chromosome evolution. We, therefore, compared rates of evolution for X-linked and autosomal coding sequence, and recovered a significant pattern of Faster-X in the guppy. X-linked d_N/d_S is greater though marginally non-significant for X-linked genes (86 genes, permutation test with 1,000 replicates, $P = 0.067$) relative to the autosomes (4,755 genes), due to a marginally significant increase in d_N (permutation test with 1000 replicates, $P = 0.014$) (Supplementary Table 6, Supplementary Fig. 5). This pattern is evident across both Strata I and II, indicating that low levels of sex chromosome divergence are sufficient to facilitate Faster-X processes.

Population variation in male colour and sexual conflict. Predation pressures vary substantially for natural guppy populations, with generally lower predation pressures upstream compared with downstream[26]. This has led to differences in female preference for male colouration[11], with downstream males less vivid due to reduced female preferences and higher predation risks than upstream populations[24,25,38]. Upstream and downstream populations within watersheds are more closely related to each other than across watersheds (Supplementary Fig. 6A). Therefore, shifts in male colouration have occurred independently in each watershed[39], where downstream males are less colourful than upstream males.

Given the very recent origin of the guppy sex chromosomes, we might expect that if recombination suppression is indeed driven by sexual conflict over colour, there might be differences in the divergence of the sex chromosomes across different populations with more or less male colouration. In line with this prediction, there is evidence that different populations of wild guppies display different patterns of Y-linkage of colour traits[40]. We, therefore, examined patterns of sex-specific heterozygosity for upstream and downstream populations of wild guppies. We sampled three watersheds (Yarra, Quare, Aripo) and from each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Our results (Fig. 4, Supplementary Table 7) show that across replicate upstream populations, where males are more colourful, there is significantly greater divergence between the X and Y chromosomes than the ancestral downstream populations (Wilcoxon rank sum test between upstream and downstream populations across watersheds, Yarra $P = 0.011$, Quare $P = 0.046$, Aripo $P = 0.017$). Expansion of the non-recombining region and corresponding X-Y divergence has occurred repeatedly and independently across populations, as the phylogeny of these populations reveals that in each watershed,
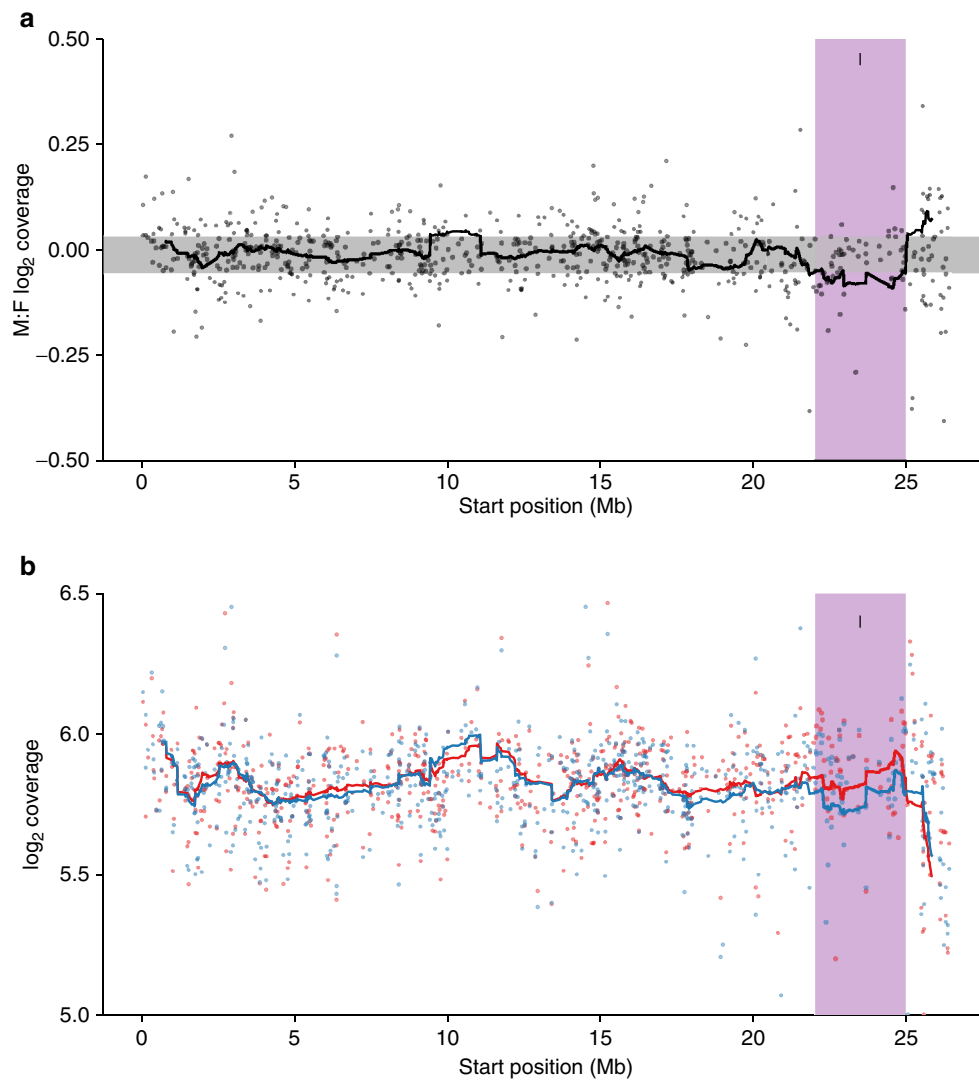
Figure 2 | Male and female coverage characteristics of guppy sex chromosome. (**a**) Moving average of coverage differences between male and female reads based on sliding window analysis (window size of 40 scaffolds). Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. (**b**) Male (blue) and female (red) coverage for the X chromosome. For both panels, dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22-25 Mb).

upstream populations are consistently derived from downstream populations (Supplementary Fig. 6B). By randomly sampling 10 Mb windows with 1,000 repetitions across the autosomes, we find that the probability of observing this convergence in SNP density across populations by chance is $P < 0.004$. In contrast, there are no differences in patterns of coverage between upstream and downstream populations in the area of greatest sex chromosome divergence (Stratum I, 22–25 Mb, Supplementary Fig. 7, Supplementary Table 8), indicating that X-Y divergence in this region predates the divergence of these wild populations.

Discussion
Observations of Y-linkage for a large proportion of male colour patterns in guppies[8,9] helped form the nucleus of theories regarding the role of sexual conflict in sex chromosome formation[3–5]. Here we used individuals from natural and laboratory populations in conjunction with analysis of coverage, SNP and expression differences between males and females in this model system to test the role of sexual conflict in recombination

suppression between the X and Y chromosomes. Our results suggest two regions of divergence on the sex chromosome. One region, likely the area of greatest Y chromosome divergence, is manifest with slightly reduced DNA coverage in males in a restricted region spanning 22–25 Mb. A larger region of more recent recombination suppression from 15–22 Mb is distinguishable only through the build-up of Y-specific SNPs. In both regions, although male-specific SNPs have accumulated on the Y, there is no evidence of large-scale decay of the Y chromosome or loss of gene activity observed in many older sex chromosome systems[31,41]. Surprisingly, this region of divergence extends over nearly half of the sex chromosome, indicating that recombination has been suppressed over a large region very recently. The two strata we observe in guppies are consistent with step-wise patterns of sex chromosome formation observed in many other organisms, including mammals[42], birds[21], *Silene*[43], sticklebacks[44] and *Nothobranchius*[45]. In the latter case, the authors observed population-level variation in the youngest stratum, similar to what we observe in guppies, suggesting that strata can form independently within species.
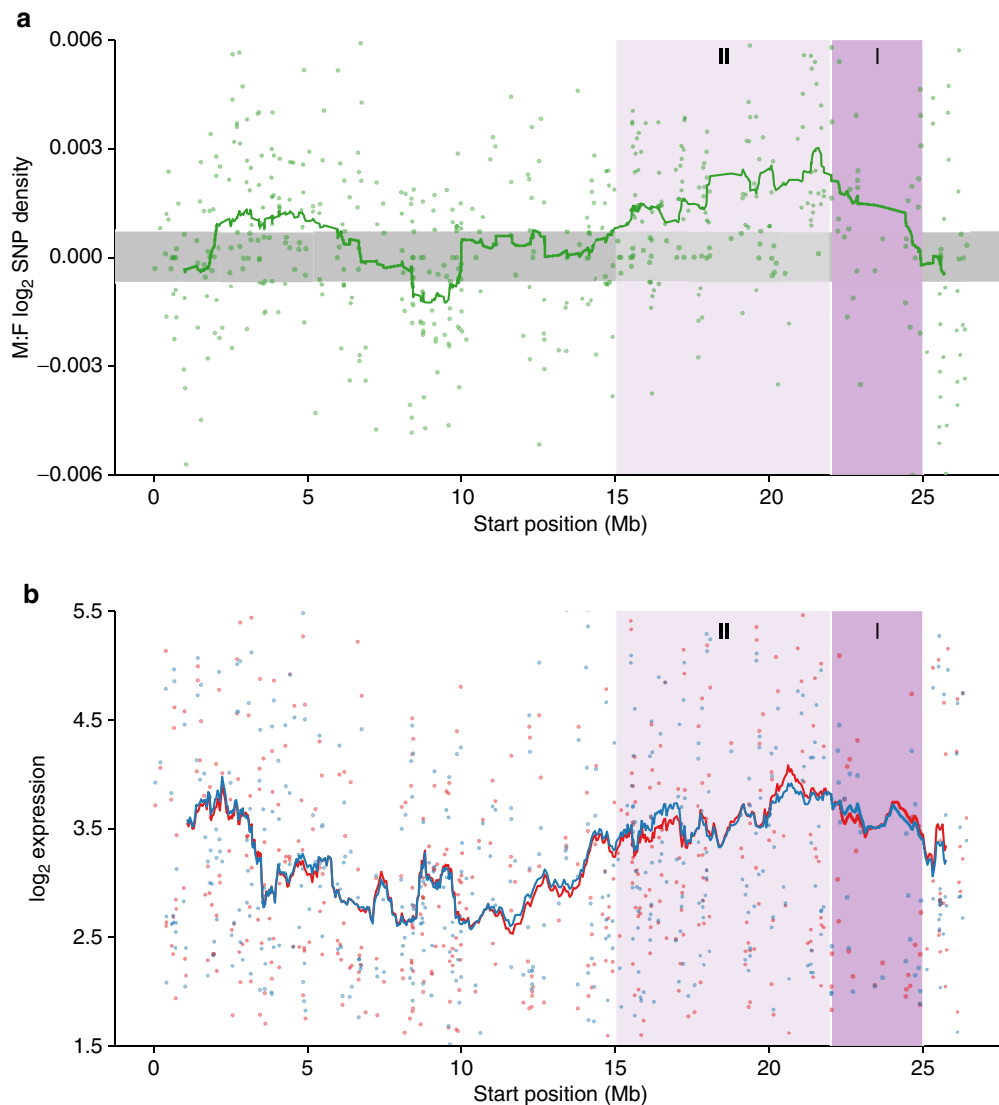
Figure 3 | Male and female SNP density and expression differences on guppy sex chromosome. (**a**) Moving average of male:female SNP density based on sliding window analysis (window size of 40 scaffolds). Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. (**b**) Male (blue) and female (red) expression of genes along the X chromosome (window size of 40 genes). Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22–25 Mb) (see Fig. 2), light purple indicates the region with less X-Y differentiation, where there is a significant excess of male SNPs (Stratum II, 15–22 Mb).

Comparisons of coverage and SNP density between males and females, like the analyses we implement here, offer two complementary views of sex chromosome evolution. Coverage differences are expected in more diverged regions with significant Y chromosome degeneration. In contrast, sex-differences in SNP density, particularly in regions with elevated SNP density in the heterogametic sex, are expected in more diverged systems with little Y chromosome degeneration. However, implementing these approaches in young sex chromosome systems should be accompanied by information as to the location of the sex determining region, which has been previously mapped to the far end of chromosome 12 (ref. 13). Ideally, Y-specific sequence data would be useful in verifying and dating stratum boundaries. However, this is complicated in our system due to the lack of complete lineage sorting of Y-specific SNPs, precluding the reconstruction of Y-specific sequence from our short-read data. In future work, long read RNA-seq data, optical mapping and other phasing approaches will be useful in confirming stratum

boundaries and identifying Y-linked sequences. These data will also be important in determining whether inversions, which are often assumed to be involved in recombination suppression, are indeed the mechanism behind sex chromosome divergence.

Despite the limited sequence divergence between the X and Y chromosomes, we observe two evolutionary signatures that are typically only associated with heteromorphic sex chromosome systems. First, the X chromosome shows the early stages of sexualization for gene expression despite limited evidence for degeneration in gene activity or content across the non-recombining Y chromosome. Previous evidence of sexualization comes from old, highly heteromorphic sex chromosome systems[46–48] and it was previously unclear how quickly sex-biased expression can accumulate after sex chromosome formation. Our results, therefore, indicate that sexualization of the X chromosome can occur very quickly after recombination is halted. Second, we detect a Faster-X effect, where X-linked coding sequence diverges more rapidly than the remainder of the
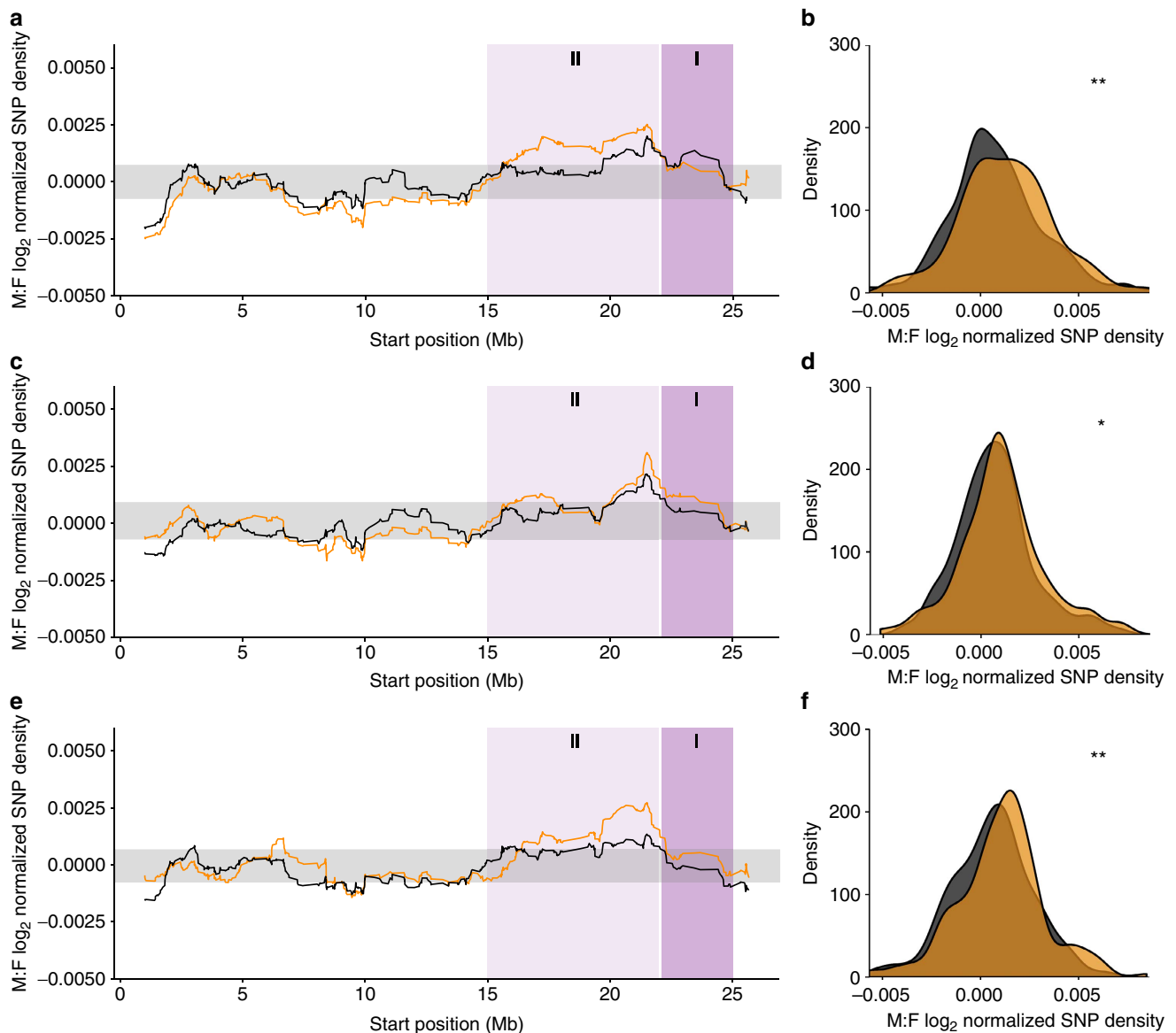
Figure 4 | Male:female SNP density for the X chromosome across upstream (orange) and downstream (black) guppy populations. (**a**,**c** and **e**) Moving averages of normalized SNP density across the X chromosome based on sliding window analysis (window size of 40 genes) for Yarra (**a**), Quare (**c**) and Aripo (**e**) watersheds. Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in laboratory population males (Stratum I, 22–25 Mb) (see Fig. 2), light purple indicates the region with less X-Y differentiation, where there is a significant excess of male SNPs in laboratory populations (Stratum II, 15–22 Mb) (see Fig. 3). (**b**,**d** and **f**) Distribution of sex differences in normalized SNP density for the X-Y diverged region (Strata I and II, 15–25 Mb) for Yarra (**b**), Quare (**d**) and Aripo (**f**) watersheds. **P-value < 0.020, *P-value < 0.050 based on permutation tests.

genome. Until now, evidence for Faster-X was restricted to highly diverged sex chromosomes[35–37], however, our results suggest that the Faster-X processes can accumulate rapidly following the loss of recombination. These findings have important consequences for the role of sex chromosomes in Haldane's rule[49] and the Large-X effect in speciation[50], and suggests that young or undifferentiated sex chromosomes may act as an important driver in the evolution of reproductive isolation[51].

Most systems where sex chromosomes have formed recently[28–30], and even some older sex chromosome systems[20,52], show restricted recombination in only a small region. The region of divergence extends over almost half of the

sex chromosomes in the guppy, suggesting that recombination has been suppressed very quickly over a large region of the Y chromosome in guppies. This rapid spread of recombination suppression may have been driven by the presence of sexually antagonistic alleles related to male colour on the proto-sex chromosome[10–12]. The high proportion of Y-linked colour patterns in guppies[8,9] is likely the product of rapid spread of recombination suppression between the X and Y chromosomes, which would resolve sexually antagonism by limiting colour expression to males.

Fish show remarkable variation in sex determination[6,53] and rapid origin and turnover of sex chromosomes[6,54]. The tiger pufferfish has homomorphic sex chromosomes, where

the sexes differ by only a single missense SNP[28], whereas a significant proportion of the sex chromosomes in sticklebacks are non-recombining[44,55], and there has been substantial decay of gene activity on the Y chromosome. Although studies in related species are required to date the exact age of the sex chromosomes in *P. reticulata*, there is extensive sex chromosome turnover in the poeciliid clade[54,56], suggesting a recent origin of the sex chromosomes described here. This is consistent with expectations that the expansion of the Y-limited region was driven by sexual conflict over colouration, suggesting that Stratum II originated around the same period that male colouration emerged as a major component of female preference, likely <5Mya (refs 22,23).

Our results suggest that this younger region of recombination suppression has expanded convergently in upstream populations as a consequence of increased sexual selection and sexual conflict over colouration[11,25]. We found the same convergent pattern of X and Y divergence between colourful upstream populations compared with the duller ancestral downstream populations over each of three replicate watersheds (Fig. 4). Upstream populations all showed greater divergence between the X and Y based on SNP density, and the region of significant SNP divergence extends over a larger region of the sex chromosomes. This accelerated divergence in upstream populations in each of the three watersheds has likely occurred independently, as populations within watersheds are well known to be monophyletic[39]. In support of this, our phylogenetic reconstruction reveals that in each watershed, upstream populations independently evolved from ancestral downstream populations. This suggests that sexual selection and sexual conflict over colour has driven greater Y divergence, consistent with longstanding theoretical predictions about the role of sexual antagonism in sex chromosome formation[3–5]. However, it is worth noting that our replicate upstream populations show some variation in the degree of differentiation, possibly due to demographic factors such as bottlenecks and recent expansions, date of colonization, rate of dispersal and gene flow between upstream and downstream populations, and effective population size, as well as stochastic processes.

Altogether, our results suggest that sexual conflict may be responsible for the remarkably rapid recent spread of recombination suppression to encompass colouration alleles within the Y chromosome. Moreover, our data are consistent with a role of sexual selection in accelerating divergence of the Y chromosome once recombination suppression is established.

Methods

Sample collection. All samples were collected in accordance with national and institutional ethical guidelines. First, we sampled males and females from a single large, outbred laboratory population established in 1998 (ref. 57). Tail samples were homogenized and stored in RNA later before RNA preparation, the remainder of each fish was stored in ethanol before DNA preparation.

Second, wild males were caught from three watersheds (Yarra, Quare, Aripo) in the Northern Range Mountains of Trinidad in February 2015 (see ref. 58 for description of the habitats). From each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Samples were collected and stored immediately in ethanol prior to DNA preparation.

Sequencing. Nucleic acids were extracted with RNAeasy Kit (Qiagen) and DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer protocols. The libraries were prepared and barcoded at The Wellcome Trust Centre for Human Genetics, University of Oxford using standard protocols. RNA was sequenced on an Illumina HiSeq 2500 resulting in on average 32 million 100 bp paired-end reads per sample. DNA was sequenced on an Illumina HiSeq 4000, resulting in on average 269 million 100 bp paired-end reads per individual sampled from a single large, outbred laboratory population, and 123 million 100 bp paired-end reads per sample for individuals caught in the wild in Trinidad (Supplementary Table 1).

Quality trimming and filtering. DNA data were quality assessed using FastQC v0.11.4 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and quality trimmed using Trimmomatic v0.35 (ref. 59). We filtered reads containing adaptor sequences and trimmed reads if the sliding window average Phred score over four bases was <15 or if the leading/trailing bases had a Phred score <3. Reads were removed post filtering if either read pair was <50 bases in length. RNA-seq data was quality assessed and trimmed using the same criteria but with a minimum length threshold of 36 bases (Supplementary Table 1).

De novo genome assembly. Reads used to construct *de novo* genome assemblies were error corrected with Quake v0.3.5, specifying default settings and a kmer length of 19 (ref. 60) (Supplementary Table 2). Optimal kmer length for *de novo* genome assemblies was estimated using kmergenie v1.6741 (ref. 61).

We constructed a female de novo genome assembly with DNA-seq reads from two females using SOAPdenovo v2.04 (ref. 62) and specifying the multi-kmer option with a starting kmer of 37 and max kmer of 55. All reads were used during both contig and scaffold assembly. During scaffolding (SOAPdenovo scaff), the –F parameter was set to specify that gaps in scaffolds should be filled. Lastly, GapCloser was used to close gaps emerging during the scaffolding process. Sequences <1 Kb in length were filtered from the assembly (Supplementary Table 3).

Assigning chromosomal position. Guppy genes were downloaded from RefSeq (Guppy_female_1.0 + MT, RefSeq assembly accession: GCF_000633615.1) and the longest isoform picked for each. Coding sequences were BLASTed against the *de novo* genome assembly using BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score.

De novo scaffolds were assigned to the guppy reference chromosomes and oriented using the chromosomal location and start position of mapped guppy genes. If multiple genes mapped to a given scaffold, the scaffold was assigned to the reference chromosome that the majority of genes were located on. Specifically, at least 70% of genes mapping to a given scaffold must be located on the same chromosome in the reference genome otherwise the scaffold was discarded. The degree of concordance in assigned chromosome position using this approach is high (Supplementary Table 4), and only 320 scaffolds from the female genome assembly were discarded due to discordance between chromosomal locations.

Genomic coverage analysis. Male and female trimmed DNA-seq reads were separately mapped to the *de novo* genome assembly using Bwa v0.7.12 aln/sampe with default settings[64]. Uniquely mapped reads were extracted using grep 'XT:A:U' and soap.coverage v2.7.9 (http://soap.genomics.org.cn) was used to extract coverage of scaffolds in every individual. For each scaffold, coverage was defined as the total number of times each site was sequenced divided by the number of sites that were sequenced.

For lab populations, average coverage values were calculated for females and males separately. We added 1 to each value to avoid infinitely high numbers associated with $\log_2$ 0. Male:female coverage was calculated for each scaffold as $\log_2$(average male coverage) – $\log_2$(average female coverage).

For upstream and downstream wild populations, coverage was estimated using Bwa v0.7.15 aln/sampe and the same pipeline as the lab populations. Average coverage was calculated for each gene separately across each population. To account for differences in sequencing depth across populations, the $\log_2$ coverage for each gene was normalized by the median $\log_2$ coverage of X chromosome ($\log_2$ coverage—median $\log_2$ coverage of X chromosome). Male:female coverage was estimated for each population relative to the normalized coverage of the female lab population.

Polymorphism analysis. Male and female trimmed DNA-seq reads from both wild and lab populations were separately mapped to the *de novo* genome assembly using Bowtie1 v1.1.2 (ref. 65), specifying a maximum insert size for paired-end alignment of 1,400 and writing hits in map format. Map files were sorted by scaffold and bow2pro v0.1 (http://guanine.evolbio.mpg.de/) was used to generate a profile for each sample. Sites with coverage <10 were excluded from the analysis and SNPs were called when a site had a major allele frequency of 0.3 times the site coverage. SNPs were only included in further analyses if they were located within genic regions (see Expression analysis method for detail on gene annotation). Average SNP density for each gene was calculated as sum(SNPs) divided by sum(no. of filtered sites). We added 1 to each value to avoid infinitely high numbers associated with $\log_2$. Genes were excluded if zero sites remained after filtering.

For lab populations, average SNP density was calculated separately for males and females. Male:female SNP density was calculated for each gene as $\log_2$(average male SNP density) − $\log_2$(average female SNP density).

For upstream and downstream wild populations, average SNP density was calculated for each gene separately across each population. To account for differences in overall genetic diversity across populations, the $\log_2$ SNP density for each gene was normalized by the median $\log_2$ SNP density of X chromosome ($\log_2$ SNP density—median $\log_2$ SNP density of X chromosome). Male:female

SNP density was estimated for each population relative to the normalized SNP density of the female lab population.

To calculate the probability that the convergence in patterns of SNP density across populations we observe is due to chance, we randomly sampled 10 Mb windows across the autosomes 1,000 times. For each window, we tested whether the upstream normalized male:female SNP density was greater than the downstream population in each river using a one-tailed Wilcoxon ranked sum test. We looked for windows where all three rivers had P-values < 0.05 and the median SNP density in the lab population was greater than the 95% autosomal confidence interval.

Expression analysis. Male and female trimmed RNA-seq reads were separately mapped to the *de novo* genome assembly using HISAT2 v2.0.4 (ref. 66), suppressing unpaired and discordant alignments for paired reads and excluding reads from the sam output that failed to align. Reported alignments were tailored for transcript assemblers including StringTie.

Sam files were coordinate sorted using SAMtools v1.2 (ref. 67) and converted to bam files. StringTie v1.2.3 (ref. 67) was used to quantify gene expression and annotate the *de novo* assembly.

Specifically, StringTie was run on each sample with default settings and the output GTF files were merged. The combined GTF file was filtered to remove non-coding RNA (ncRNA) and transcripts less than 50 bp in length. Specifically, transcript sequences were extracted using bedtools getfasta[68] and BLASTed to *Oryzias latipes* (MEDAKA1), *Gasterosteus aculeatus* (BROADS1), *Poecilia formosa* (PoeFor_5.1.2) and *Danio rerio* (GRCz10) ncRNA downloaded from Ensembl 84 (ref. 69). Transcripts with blast hits to ncRNA were removed from the GTF file. StringTie was rerun on each sample and expression was only estimated for genes defined in the filtered GTF file. A minimum expression threshold of 2FPKM in at least half of the individuals of either sex was imposed. This final filtered data set (23,603 genes) was used in subsequent expression and polymorphism analyses.

Expression was normalized using EdgeR[70]. Sam files were name sorted using SAMtools and HT-seq count v0.6.1 (ref. 71) used to extract read counts for each gene. Genes were excluded if they were not located on scaffolds assigned to the guppy reference genome. In all, 13,306 genes remained after filtering. Expression was normalized using TMM (trimmed mean of m-values) in EdgeR and RPKM estimated for each gene. Individuals cluster transcriptomically by sex (Supplementary Fig. 4). Average RPKM for each gene was calculated separately for males and females. We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Male:female expression was calculated for each gene as $\log_2$ (average male RPKM)—$\log_2$(average female RPKM).

We tested whether there was an enrichment of GO terms in the X-Y diverged region compared with the rest of the genome. *Danio rerio* (GRCz10) coding sequences were downloaded from Ensembl 84 (ref. 69) and the longest isoform extracted for each gene. Longest isoforms were extracted for our set of expressed guppy genes and BLASTed to *D. rerio* using BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score. *D. rerio* orthologues were identified for genes in the X-Y degenerate region (15–25 Mb) and compared with the remainder of the genome using GOrilla[72,73].

Cluster analysis of expression data. Transcriptional similarity of normalized count data for female and male individuals was assessed using a multi-dimensional scaling plot (MDS) with default settings in EdgeR[74]. RPKM data was clustered using the R package pheatmap and boostrap values calculated using pvclust. UPGMA was used in the hierarchical cluster analysis and the distance matrix was computed using the Euclidean method.

Calculating moving averages. Moving averages of coverage/polymorphism/ expression were calculated in R[75] based on sliding window analyses using the roll_mean function. Ninety-five per cent confidence intervals for the moving average were calculated by randomly resampling (1,000 times, without replacement) autosomal scaffolds (coverage analysis) or genes (SNP density and expression analyses).

Faster-X analysis. Guppy transcript sequences were extracted using bedtools getfasta[68] and the longest isoform chosen for each of the 23,603 genes. Genes on genomic scaffolds without chromosomal locations were removed, leaving 13,306 genic sequences for the Faster-X analysis. *Oryzias latipes* (MEDAKA1), *Xiphophorus maculatus* (Xipmac4.4.2), *Poecilia formosa* (PoeFor_5.1.2) were downloaded from Ensembl 84 (ref. 69) and the longest transcript for each gene was identified. We determined orthology using reciprocal BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score. In all, 7,382 reciprocal 1-1 orthologues across the four species were identified. We obtained open reading frames and protein coding sequence with BLASTx v2.3.0 with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30% using the approach in Wright et al.[76] Reciprocal orthologues with no BLASTx hits or a valid protein-coding sequence were excluded.

Reciprocal orthologues were aligned with PRANK v.140603 (ref. 77) using the codon model and specifying the following guidetree; (((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*), *Oryzias latipes*). SWAMP v 31-03-14 (ref. 78) was used to mask erroneous sequences in the alignments. Reciprocal orthologues were discarded if the alignment length was < 300 bp after removing gaps and masked sites. After this length filter, 5,349 reciprocal orthologues remained.

We used the branch model (model = 2, nssites = 0) in the CODEML package in PAML v4.8 (ref. 79) to obtain divergence estimates using the following phylogeny; ((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*, *Oryzias latipes*). The branch model was used to calculate mean d_N/d_S across the *Poecilia reticulata* branch. As mutational saturation and double hits can lead to inaccurate divergence estimates[80] orthogroups were excluded if $d_S > 2$.

Orthologues were divided into genomic categories on the basis of their chromosomal location. For each category, mean d_N and mean d_S were calculated as the sum of the number of substitutions across all orthologues divided by the number of sites ($d_N = $ sum D_N/sum N, $d_S = $ sum D_S/sum S, where D_N and D_S are estimates of the number of nonsynonymous or synonymous substitutions and N and S are the number of nonsynonymous/synonymous sites). This approach prevents disproportionate weighting of shorter genes by avoiding the problems of infinitely high d_N/d_S estimates arising from sequences with extremely low d_S (refs 76,81,82).

Significant differences in d_N, d_S and d_N/d_S between genomic categories were determined using permutation tests with 1,000 replicates. One-tailed tests were used to test for the Faster-X effect where we predict d_N/d_S is greater for X-linked gene relative to the autosomes. Two-tailed tests were used to test for differences in d_N and d_S. Bootstrapping with 1,000 repetitions was used to generate 95% confidence intervals.

Phylogenetic history of guppy populations. Using DNA-seq data, we reconstructed the phylogenetic relationships between the six wild populations. We mapped trimmed reads to the previously sequenced guppy genome (Guppy_female_1.0 + MT, RefSeq assembly accession: GCF_000633615.1) using Stampy v1.0.28 (ref. 83) with a substitution rate of 0.01. After mapping, sam files were converted to bam and coordinate sorted using SAMtools v1.2 (ref. 84) and then deduplicated using Picard tools v1.136 (ref. 85). Subsequently, we added read groups and merged libraries belonging to the same individual using Picard. We then called variants on all 24 individuals simultaneously using two independent methods (GATK and Platypus), and retained only SNPs called reliably with both methods and passing quality control filters.

As part of the GATK variant calling pipeline, v3.4.46 (ref. 86), we first realigned reads around indels and recalibrated base quality scores. We then proceeded with variant calling using the HaplotypeCaller and GenotypeGVCFs tools. The second method we employed to call variants was Platypus v0.8.1 (ref. 87), which we ran in assembly mode, restricting calling to reads mapping to the 23 canonical chromosomes (that is, excluding those mapped to unplaced scaffolds).

After variant calling we removed indels, intersected the GATK and Platypus SNP sets, and applied stringent quality filtering. We removed singleton SNPs, multiallelic SNPs and SNPs failing the following quality thresholds: quality by depth > 2, coverage > 0.5x and < 2x mean coverage, > 2 reads for the alternative allele, mapping quality > 40, allele bias Z score for mapping quality, base quality or read position < -1.96, or strand bias Fisher exact test $P > 0.05$. We also removed SNPs with missing genotype in any individual. This yielded 4.6 million high-quality SNPs.

Next, we used R package adegenet v2.0.1 (ref. 88) to construct a Euclidian distance matrix for the 24 individuals based either on all SNPs across the genome or on only the 72,623 SNPs between 15 and 25 MB on the X chromosome. We used the R package ape v3.5 (ref. 89) to produce from each matrix a simple neighbour joining tree to visualize the genetic distance between the six populations, and performed 100 bootstrap iterations to assess support for each node.

References

1. Bachtrog, D. Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. *Nat. Rev. Genet.* **14,** 113–124 (2013).
2. Bachtrog, D. *et al.* Are all sex chromosomes created equal? *Trends Genet.* **27,** 350–357 (2011).
3. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, 1983).
4. Fisher, R. A. The evolution of dominance. *Biol. Rev.* **6,** 345–368 (1931).
5. Rice, W. R. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41,** 911–914 (1987).
6. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol.* **12,** e1001899 (2014).
7. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7,** 12087 (2016).

8. Lindholm, A. & Breden, F. Sex chromosomes and sexual selection in poeciliid fishes. *Am. Nat.* **160,** S214–S224 (2002).

9. Winge, M. The location of eighteen genes in *Lebistes reticulata. J. Genet.* **18,** 1–43 (1927).

10. Endler, J. A. Natural selection on color patterns in Poecilia reticulata. *Evolution* **34,** 76–91 (1980).

11. Houde, A. E. & Endler, J. A. Correlated evolution of female mating preferences and male color patterns in the guppy, Poecilia reticulata. *Science* **248,** 1405–1408 (1990).

12. Kemp, D. J., Reznick, D. N., Grether, G. F. & Endler, J. A. Predicting the direction of ornament evolution in Trinidadian guppies (Poecilia reticulata). *Proc. R. Soc. B-Biol. Sci.* **276,** 4335–4343 (2009).

13. Tripathi, N., Hoffmann, M., Weigel, D. & Dreyer, C. 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 (2009).

14. Lisachov, A. P., Zadesenets, K. S., Rubtsov, N. B. & Borodin, P. M. Sex chromosome synapsis and recombination in male guppies. *Zebrafish* **12,** 174–180 (2015).

15. Lenormand, T. The evolution of sex dimorphism in recombination. *Genetics* **163,** 811–822 (2003).

16. Nanda, I. *et al.* Sex chromosome polymorphism in guppies. *Chromosoma* **123,** 373–383 (2014).

17. Traut, W. & Winking, H. Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res.* **9,** 659–672 (2001).

18. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825–U822 (2003).

19. Vicoso, B., Emerson, J. J., Zektser, Y., Manajan, S. & Bachtrog, D. Comparative sex chromosome divergence in snakes: differentiation and lack of global dosage compensation. *PLoS Biol.* **11,** e1001643 (2013).

20. Vicoso, B., Kaiser, V. B. & Bachtrog, D. Sex-biased gene expression at homomorphic sex chromosomes in Emus and its implications for sex chromosome evolution. *Proc. Natl Acad. Sci., USA* **110,** 6453–6458 (2013).

21. Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A. & Mank, J. E. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* **68,** 3281–3295 (2014).

22. Meredith, R. W., Pires, M. N., Reznick, D. N. & Springer, M. S. Molecular phylogenetic relationships and the evolution of the placenta in Poecilia (Micropoecilia) (Poeciliidae: Cyprinodontiformes). *Mol. Phylogenet. Evol.* **55,** 631–639 (2010).

23. Pollux, B. J. A., Meredith, R. W., Springer, M. S., Garland, T. & Reznick, D. N. The evolution of the placenta drives shift in sexual selection in livebearing fish. *Nature* **513,** 233–236 (2014).

24. Endler, J. A. Nateral selection on color patterns in Poecilia reticulata. *Evolution* **34,** 76–91 (1980).

25. Endler, J. A. Natural and sexual selection on color patterns in Poeciliid fishes. *Env. Biol. Fish.* **9,** 173–190 (1983).

26. Endler, J. A. Multiple trait coevolution and environmetnal gradients in guppies. *Trends Ecol. Evol.* **10,** 22–29 (1995).

27. Gordon, S. P., Lopez-Sepulcre, A. & Reznick, D. N. Predation-associated differences in sex linkage of wild guppy coloration. *Evolution* **66,** 912–918 (2012).

28. Kamiya, T. *et al.* A Trans-Species missense SNP in Amhr2 Is Associated with Sex determination in the Tiger Pufferfish, Takifugu rubripes (Fugu). *Plos Genet.* **8,** e1002798 (2012).

29. Liu, Z. Y. *et al.* A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427,** 348–352 (2004).

30. Russell, J. R. W. & Pannell, J. R. Sex determination in dioecious Mercurialis annua and its close diploid and polyploid relatives. *Heredity* **114,** 262–271 (2015).

31. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *Plos Biol.* **13,** e1002078 (2015).

32. Hough, J., Hollister, J. D., Wang, W., Barrett, S. C. H. & Wright, S. I. Genetic degeneration of old and young Y chromosomes in the flowering plant Rumex hastatulus. *Proc. Natl Acad. Sci. USA* **111,** 7713–7718 (2014).

33. Muyle, A. *et al.* Rapid de novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. *Plos Biol.* **10** (2012).

34. Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex-chromosomes and autosomes. *Am. Nat.* **130,** 113–146 (1987).

35. Zhou, Q. & Bachtrog, D. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila. Science* **337,** 341–345 (2012).

36. Meisel, R. P. & Connallon, T. The faster-X effect: integrating theory and data. *Trends Genet.* **29,** 537–544 (2013).

37. Mank, J. E., Vicoso, B., Berlin, S. & Charlesworth, B. Effective population size and the faster-x effect: empirical results and their interpretation. *Evolution* **64,** 663–674 (2010).

38. Houde, A. E. & Endler, J. A. Correlate evolution of female mating preferences and male color patterns in the guppy Poecilia reticulata. *Science* **248,** 1405–1408 (1990).

39. Fraser, B. A., Kunstner, A., Reznick, D. N., Dreyer, C. & Weigel, D. Population genomics of natural and experimental populations of guppies (Poecilia reticulata). *Mol. Ecol.* **24,** 389–408 (2015).

40. Gordon, S. P. *et al.* Selection analysis on the rapid evolution of a secondary sexual trait. *Proc. R. Soc. B-Biol. Sci.* **282,** 20151244 (2015).

41. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *Plos Biol.* **13,** 22 (2015).

42. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825–837 (2003).

43. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia:* Evidence from new sex-linked genes. *Genetics* **175,** 1945–1954 (2007).

44. White, M., Kitano, J. & Peichel, C. L. Purifying selection maintains dosage sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol. Biol. Evol.* **32,** 1981–1995 (2015).

45. Reichwald, K. *et al.* Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* **163,** 1527–1538 (2015).

46. Arunkumar, K. P., Mita, K. & Nagaraju, J. The silkworm Z chromosome is enriched in testis-specific genes. *Genetics* **182,** 493–501 (2009).

47. Meisel, R. P., Malone, J. H. & Clark, A. G. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res.* **22,** 1255–1265 (2012).

48. Wright, A. E., Moghadam, H. K. & Mank, J. E. Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics* **192,** 1433-＋ (2012).

49. Haldane, J. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12,** 101–109 (1922).

50. Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *Plos Biol.* **5,** 1890–1898 (2007).

51. Dufresnes, C. *et al.* Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. *Sci. Rep.* **6,** 21029 (2016).

52. Stock, M. *et al.* Ever-young sex chromosomes in European tree frogs. *Plos Biol.* **9,** e1001062 (2011).

53. Mank, J. E., Promislow, D. E. L. & Avise, J. C. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87,** 83–93 (2006).

54. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208,** 191–364 (2002).

55. Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461,** 1079–1083 (2009).

56. Mank, J. E., Promislow, D. E. L. & Avise, J. C. Evolution of alternative sex determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87,** 83–93 (2006).

57. Kortrschal, A. *et al.* Experimental evidence for costs and benefits of evolving a larger brain. *Curr. Biol.* **23,** 168–171 (2013).

58. Sandkam, B., Young, C. M. & Breden, F. Beauty in the eye of the beholder: colour vision is tuned to mate preference in the Trinidadian Guppy (Poecilia reticulata). *Mol. Ecol.* **24,** 596–609 (2016).

59. Lohse, M. *et al.* RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **40,** W622–W627 (2012).

60. Kelley, D. R., Schatz, M. C. & Salzberg, S. L. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* **11,** R116 (2010).

61. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30,** 31–37 (2014).

62. Luo, R. B. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1,** 18 (2012).

63. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. basic local alignment search tooL. *J. Mol. Biol.* **215,** 403–410 (1990).

64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

65. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

66. Kim, D., Landmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–U121 (2015).

67. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33,** 290–295 (2015).

68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

69. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** D749–D755 (2014).

70. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

71. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

72. Eden, E., Lipson, D., Yogev, S. & Yakhini, Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* **3,** 508–522 (2007).

73. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).

74. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 138–140 (2010).

75. R Core Team. R: A language and environment for statistical computing: R Foundation for Statistical Computing. https://www.R-project.org/ (2015).

76. Wright, A. E. *et al.* Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Mol. Ecol.* **24,** 1218–1235 (2015).

77. Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102,** 10557–10562 (2005).

78. Harrison, P. W., Jordan, G. E. & Montgomery, S. H. SWAMP: Sliding Window Alignment Masker for PAML. *Evol. Bioinform.* **10,** 197–204 (2014).

79. Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).

80. Axelsson, E. *et al.* Natural selection in avian protein-coding genes expressed in brain. *Mol. Ecol.* **17,** 3008–3017 (2008).

81. Mank, J. E., Axelsson, E. & Ellegren, H. Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res.* **17,** 618–624 (2007).

82. Mank, J. E., Nam, K. & Ellegren, H. Faster-Z evolution is predominantly due to genetic drift. *Mol. Biol. Evol.* **27,** 661–670 (2010).

83. Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).

84. Li, H. *et al.* Sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

85. Broad Institute. Picard toolkit. http://broadinstitute.github.io/picard.

86. McKenna, A. *et al.* The genome analysis toolkit: A Mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

87. Rimmer, A. *et al.* Intrating mapping, assembly and haplotype based approaches for calling variance in clinical sequencing applications. *Nat. Genet.* **46,** 912–918 (2014).

88. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27,** 3070–3071 (2011).

89. Paradis, E., Claude, J. & Strimmer, K. APE: Analysis of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–290 (2004).

Acknowledgements

Author contributions

J.E.M. conceived of the study, and J.E.M. and A.E.W. designed the experiments. N.K., S.D.B., B.S. and F.B. collected samples used for sequencing in this study. A.E., I.D., N.I.B., V.O. and B.V. analysed the data. All authors contributed to writing the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wright, A. E. *et al.* Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* **8,** 14251 doi: 10.1038/ncomms14251 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Table 1: Sequencing information for each sample

Sample	Source	Method	Coverage	Raw paired reads	Paired reads after trimming	% removed
Female_4	Lab pop	DNA-seq	79X	276,700,698	253,668,178	8.32
Female_7	Lab pop	DNA-seq	75X	262,392,382	233,869,784	10.87
Male_11	Lab pop	DNA-seq	68X	238,377,854	217,155,394	8.90
Male_8	Lab pop	DNA-seq	85X	298,121,974	271,223,610	9.02
Female_1	Lab pop	RNA-seq	-	31,614,139	30,942,700	2.12
Female_4	Lab pop	RNA-seq	-	42,162,172	41,258,427	2.14
Female_7	Lab pop	RNA-seq	-	29,059,783	28,520,214	1.86
Female_13	Lab pop	RNA-seq	-	32,392,057	31,628,760	2.36
Male_5	Lab pop	RNA-seq	-	26,029,449	25,434,019	2.29
Male_8	Lab pop	RNA-seq	-	31,955,977	31,182,426	2.42
Male_11	Lab pop	RNA-seq	-	36,271,770	35,430,424	2.32
Male_17	Lab pop	RNA-seq	-	30,579,016	29,853,906	2.37
Male_6	Lab pop	RNA-seq	-	31,303,333	30,559,027	2.38
Male_9	Lab pop	RNA-seq	-	37,546,710	36,718,647	2.21
Male_15	Lab pop	RNA-seq	-	29,875,402	29,258,994	2.06
Male_18	Lab pop	RNA-seq	-	39,603,241	38,785,961	2.06
Male_12	Lab pop	RNA-seq	-	32,568,129	31,793,985	2.38
Male_2	Lab pop	RNA-seq	-	23,196,956	22,777,671	1.81
Male_14	Lab pop	RNA-seq	-	31,859,225	31,110,939	2.35
Male_1	Wild AD	DNA-seq	37X	130,231,185	118,932,936	8.68
Male_3	Wild AD	DNA-seq	29X	101,814,846	93,738,591	7.93
Male_4	Wild AD	DNA-seq	30X	104,256,689	96,147,514	7.78
Male_5	Wild AD	DNA-seq	41X	1438,00,633	131,847,549	8.31
Male_12	Wild YU	DNA-seq	36X	126,989,240	114,251,020	10.03
Male_13	Wild YU	DNA-seq	40X	140,058,828	129,438,072	7.58
Male_14	Wild YU	DNA-seq	30X	103,948,475	95,917,845	7.73
Male_15	Wild YU	DNA-seq	29X	102,834,035	93,981,379	8.61
Male_16	Wild QD	DNA-seq	36X	126,993,935	115,367,896	9.15
Male_17	Wild QD	DNA-seq	33X	117,037,503	107,793,725	7.90
Male_19	Wild QD	DNA-seq	37X	129,116,621	118,530,011	8.20
Male_20	Wild QD	DNA-seq	30X	105,910,625	97,094,279	8.32
Male_32	Wild AU	DNA-seq	38X	133,405,487	121,078,915	9.24
Male_33	Wild AU	DNA-seq	36X	126,870,348	116,645,902	8.06
Male_34	Wild AU	DNA-seq	34X	120,555,304	111,191,193	7.77
Male_35	Wild AU	DNA-seq	33X	116,057,513	105,723,853	8.90
Male_41	Wild YD	DNA-seq	40X	139,172,741	127,519,332	8.37
Male_42	Wild YD	DNA-seq	37X	129,566,748	119,662,352	7.64
Male_43	Wild YD	DNA-seq	27X	94,884,948	87,529,008	7.75
Male_44	Wild YD	DNA-seq	30X	103,984,825	94,569,988	9.05
Male_51	Wild QU	DNA-seq	28X	98,581,145	89,880,179	8.83
Male_52	Wild QU	DNA-seq	36X	125,273,377	112,674,653	10.06
Male_53	Wild QU	DNA-seq	31X	107,398,826	99,357,018	7.49

| Male_54 | Wild QU | DNA-seq | 46X | 160,156,505 | 147,797,565 | 7.72 |

A = Aripo watershed, Y = Yarra watershed, Q = Quare watershed
U = Upstream population, D = Downstream population

Supplementary Table 2: Error corrected reads for de novo genome assembly

Sex	Source	Sequencing	No. of individuals	Coverage	Error corrected paired reads
Female	Lab pop	DNA-seq	2	137X	479,641,121

Supplementary Table 3: Female de novo genome assembly statistics

Pre 1kb length filter				Post 1kb length filter				Oriented scaffolds			
No.	N50 (Kb)	N90 (Kb)	Length (Mb)	No.	N50 (Kb)	N90 (bp)	Length (Mb)	No.	N50 (Kb)	N90 (Kb)	Length (Mb)
2,361,160	5.4	0.1	963.6	96,611	11.3	2.8	634.8	19,206	17.4	5.6	219.5

Supplementary Table 4: Assignment of chromosomal position

Reference genome	Genes	Genes mapped to assembly	Scaffolds with mapped genes	Scaffolds discarded due to mapping discordance	Oriented scaffolds	% with multiple mapped genes all from the same reference chromosome
Guppy	25,694	25,460	19,526	320	19,206	92%

Supplementary Table 5: Coverage and SNP density

	Autosomes				X chromosome			
	No.	M $\log_2$ median	F $\log_2$ median	M:F $\log_2$ median	No.	M $\log_2$ median	F $\log_2$ median	M:F $\log_2$ median
Coverage (Wilcoxon rank sum test p-value)	17,353	5.8505	5.8477	0.0024	709	5.8273 (0.0224)	5.8324 (< 0.001)	-0.0115 (<0.001)
SNP density (Wilcoxon rank sum test p-value)	13,286	0.0023	0.0022	0.0000	555	0.0030 (<0.001)	0.0023 (0.837)	0.0007 (<0.001)

Wilcoxon rank sum test between autosomal and X chromosome medians

Supplementary Table 6: Faster-X Effect

	Autosomes	X-Y diverged region	Stratum II	Stratum I
No.	4755	86	70	16
d_N/d_S	0.091	0.107	0.105	0.106
(95% CI)	0.088-0.094	0.088-0.129	0.084-0.131	0.067-0.172
(p-value)		0.067	0.107	0.249
d_N	0.003	0.004	0.004	0.006
(95% CI)	0.003-0.003	0.003-0.005	0.003-0.004	0.004-0.009
(p-value)		**0.012**	0.182	**0.002**
d_S	0.034	0.037	0.034	0.057
(95% CI)	0.033-0.034	0.033-0.042	0.030-0.039	0.043-0.069
(p-value)		0.1	0.77	**<0.001**

X-Y diverged region refers to Stratum II (15-22Mb) & Stratum I (> 22Mb)
P-values calculated relative to the autosomes using permutation tests with 1000
replicates. Confidence intervals calculated using bootstrapping with 1000 replicates.

Supplementary Table 7: Normalised SNP densities in the X-Y diverged region across upstream and downstream guppy populations

|		Downstream			Upstream	
River	No.	M $\log_2$ median	M:F $\log_2$ median	No.	M $\log_2$ median	M:F $\log_2$ median
Yarra	197	-0.0001	0.0006	194	0.0006	0.0013
(Wilcoxon rank sum test p-value)					(0.0002)	(0.0105)
Quare	196	-0.0001	0.0006	196	0.0001	0.0009
(Wilcoxon rank sum test p-value)					(0.0090)	(0.0342)
Aripo	195	-0.0000	0.0007	197	0.0002	0.0013
(Wilcoxon rank sum test p-value)					(0.0039)	(0.0209)

X-Y diverged region refers to Stratum II (15-22Mb) & Stratum I (> 22Mb)
Wilcoxon rank sum test between downstream and upstream medians

Supplementary Table 8: Normalised coverage in Stratum I of the sex chromosome across upstream and downstream guppy populations

River	Downstream			Upstream		
	No.	M $\log_2$ median	M:F $\log_2$ median	No.	M $\log_2$ median	M:F $\log_2$ median
Yarra (Wilcoxon rank sum test p-value)	59	0.0183	-0.0484	59	0.0065 (0.7345)	-0.0330 (0.8548)
Quare (Wilcoxon rank sum test p-value)	59	0.0365	-0.0289	59	0.0037 (0.5011)	-0.0444 (0.2494)
Aripo (Wilcoxon rank sum test p-value)	59	0.0292	-0.0519	59	0.0241 (0.6129)	-0.0329 (0.9314)

Wilcoxon rank sum test between downstream and upstream medians
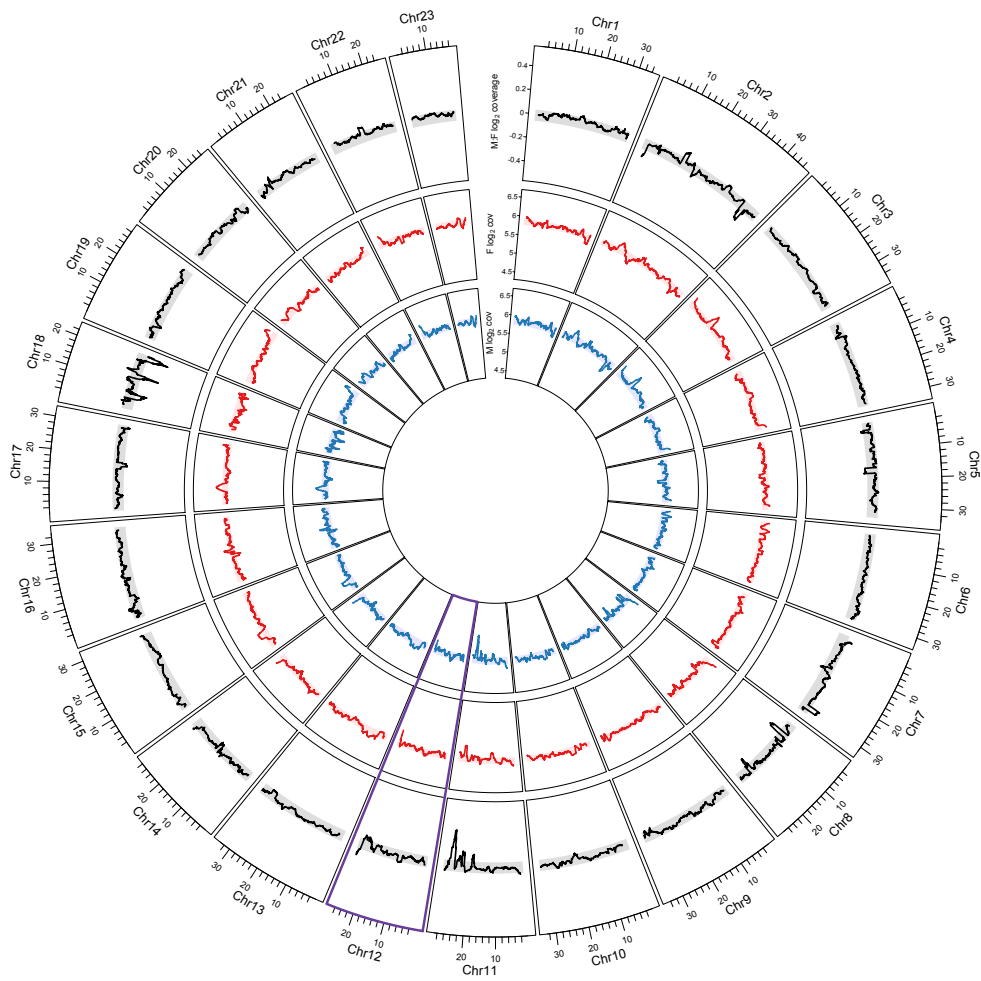
Supplementary Fig. 1. Circos plot of male and female coverage for oriented scaffolds, with moving averages based on window sizes of 40 scaffolds. For each chromosome, male coverage (blue), female coverage (red) and male:female coverage (black) is shown. 95% confidence intervals based on bootstrapping autosomal estimates are shown for male coverage (light blue), female coverage (pink) and male:female coverage (grey). The X chromosome, which contains the sex determining gene, is highlighted in purple.
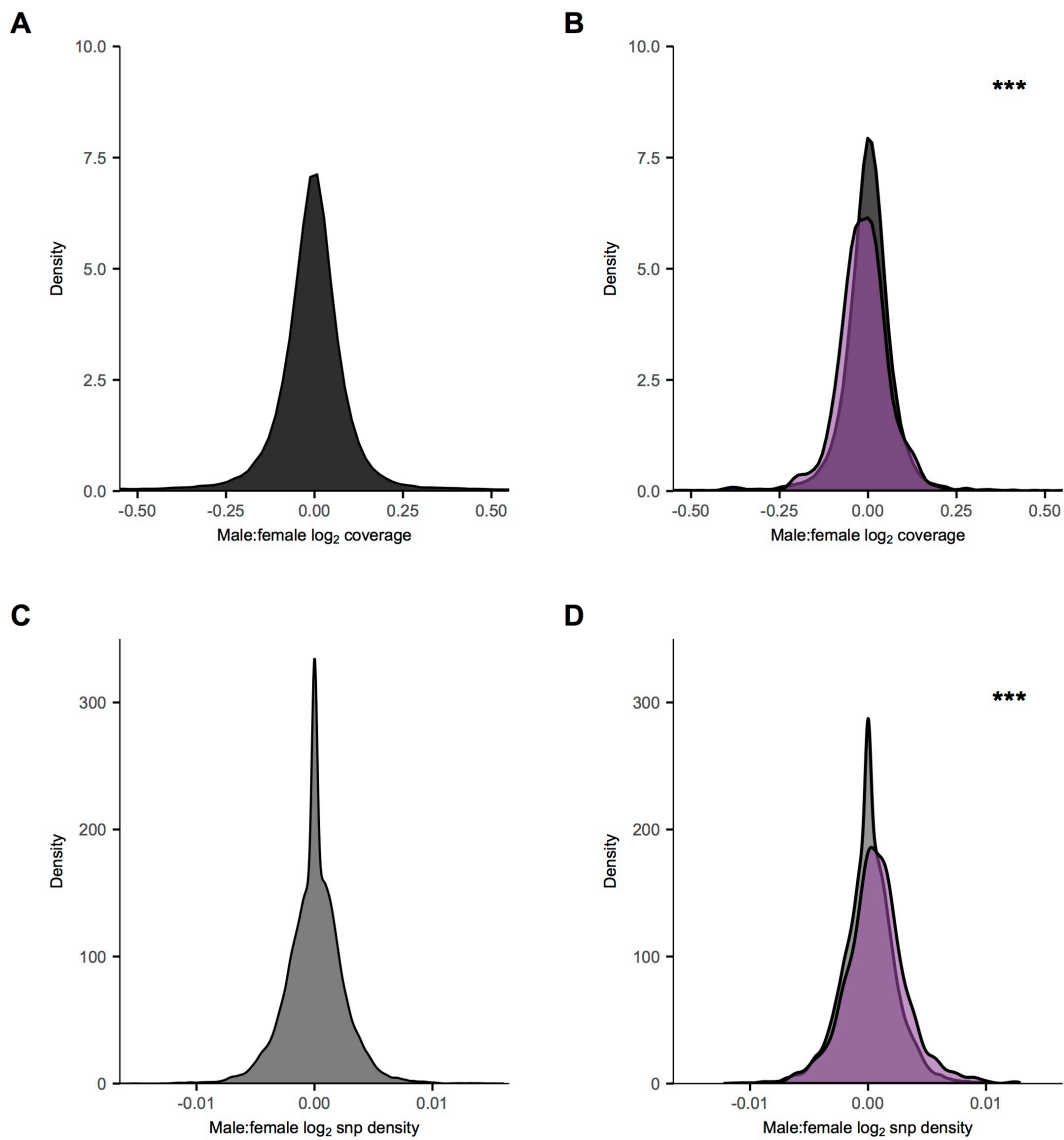
Supplementary Fig. 2. Distribution of male:female coverage (A and B) and SNP density (C and D) for assembled scaffolds. Panel A. Coverage differences for all scaffolds greater than 1Kb in length. Panel. B. Coverage differences for autosomal (grey) and the X chromosome (purple) scaffolds. Panel C. Differences in male and female SNP density for all scaffolds greater than 1Kb in length. Panel D. Differences in male and female SNP density for autosomes (grey) and the X chromosome (purple). *** p-value < 0.001.
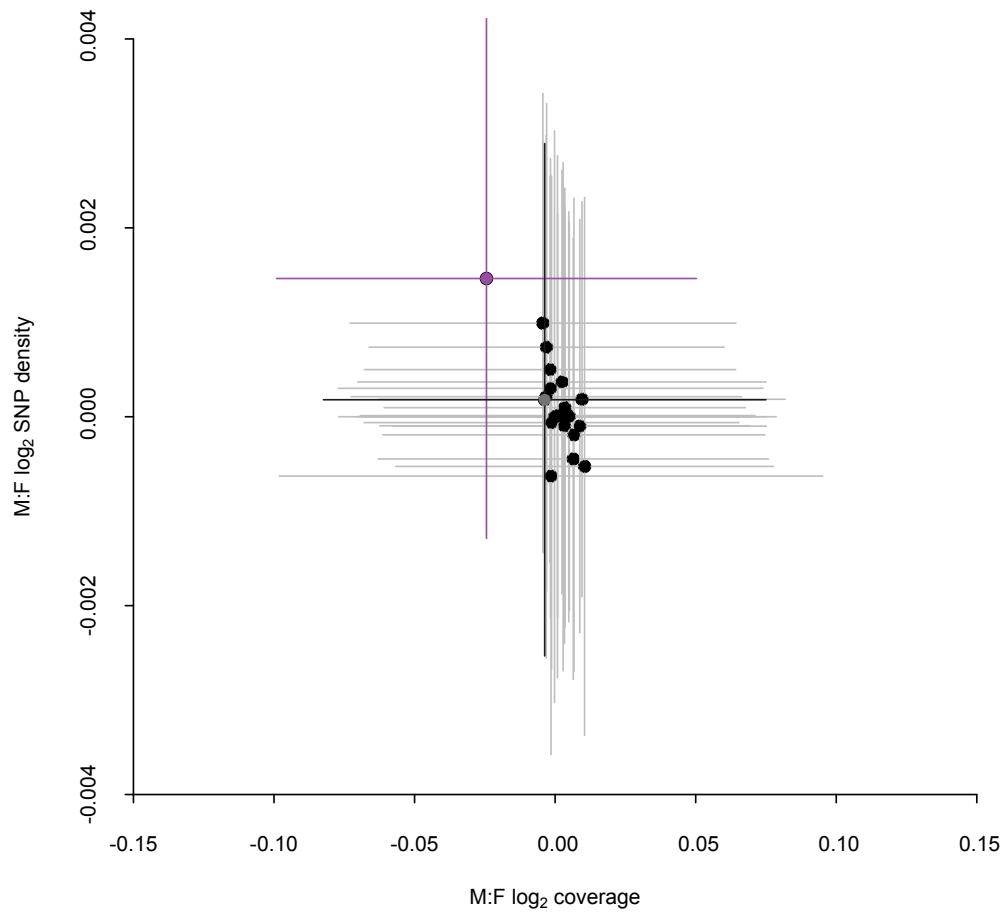
Supplementary Fig. 3. Distribution of sex differences in coverage and SNP density for all chromosomes. The X-Y diverged region (Strata I & II, 15 – 25 Mb) is in purple and PAR (<15 Mb) is in grey. Horizontal and vertical lines denote interquartile ranges.
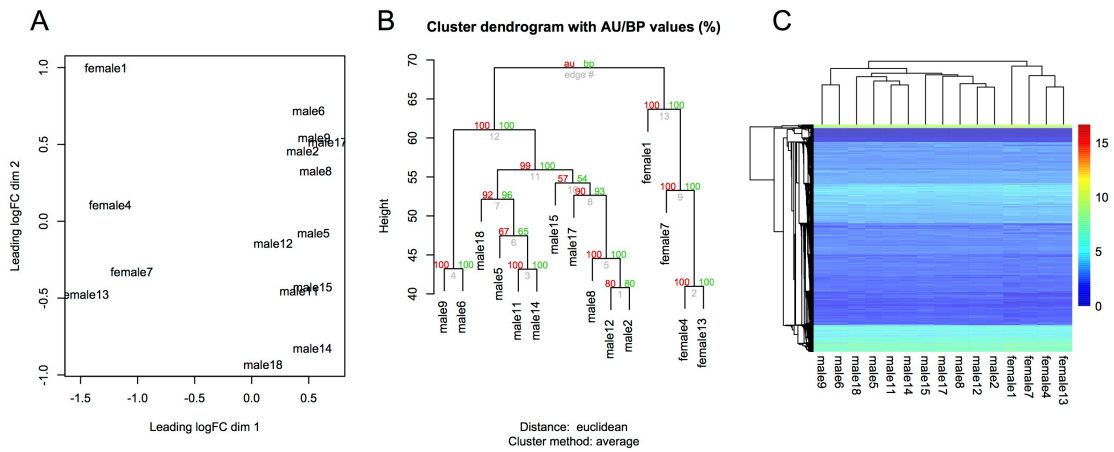
Supplementary Fig. 4. Cluster analysis of expression data. Panel A. MDS plot of normalized count data. Panel B. Cluster dendrogram of normalized log$_2$ RPKM values. Approximately unbiased p-values are shown in red and bootstrap probability values are shown in green. Panel C. Clustered heatmap of normalized log$_2$ RPKM values.
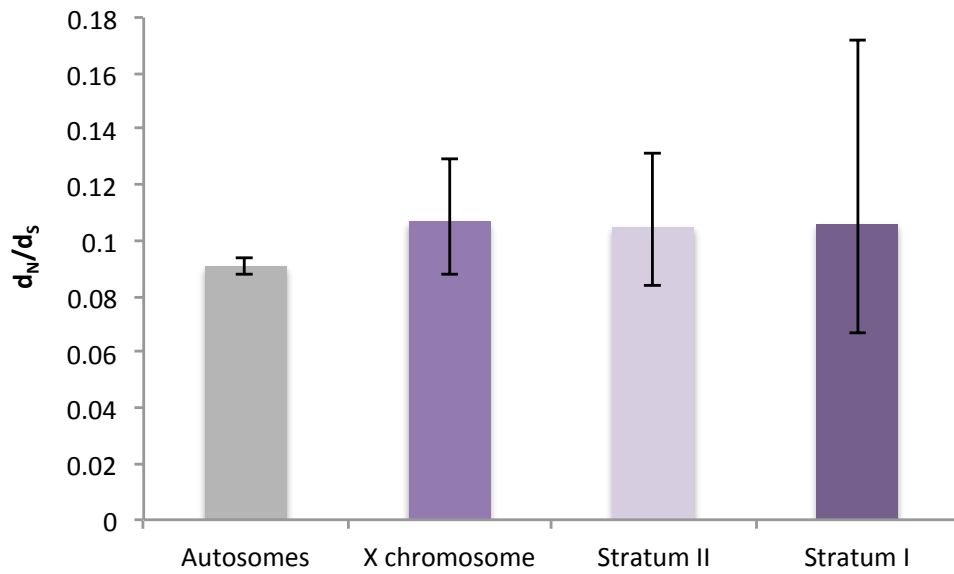
Supplementary Fig. 5. Estimates of mean d_N/d_S for the autosomes and the X chromosome. 95% confidence intervals were calculated by bootstrapping with 1000 replicates. X chromosome refers only to the X-Y region (Strata I & II, 15 – 25 Mb) and does not include genes in the PAR.
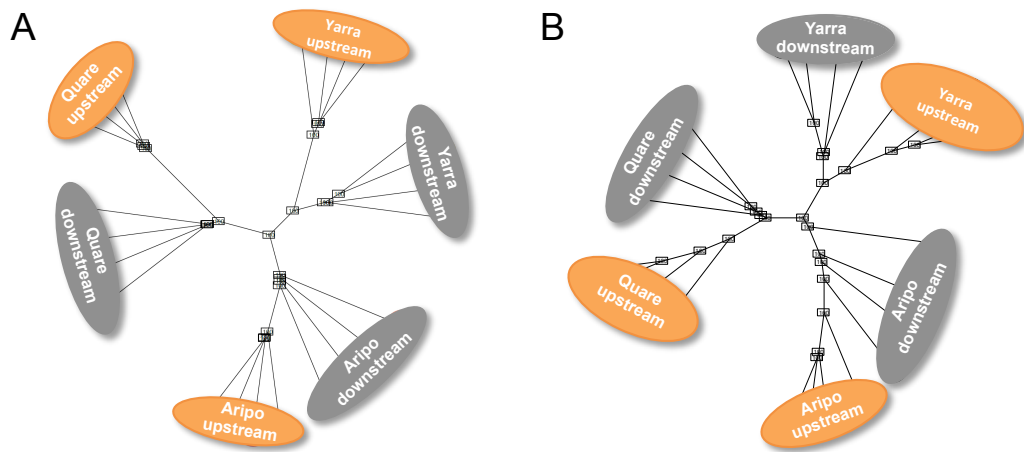
Supplementary Fig. 6. Phylogeny of upstream (orange) and downstream (grey) guppy populations across three watersheds (Yarra, Quare, Aripo) in Trinidad based on (A) all 4.6 million SNPs across the whole genome or (B) 72,623 SNPs located on the X-Y diverged region (Strata I & II, 15 – 25 Mb). Neighbor joining trees were constructed from Euclidian distances between individuals. Bootstrap support is indicated at each node.
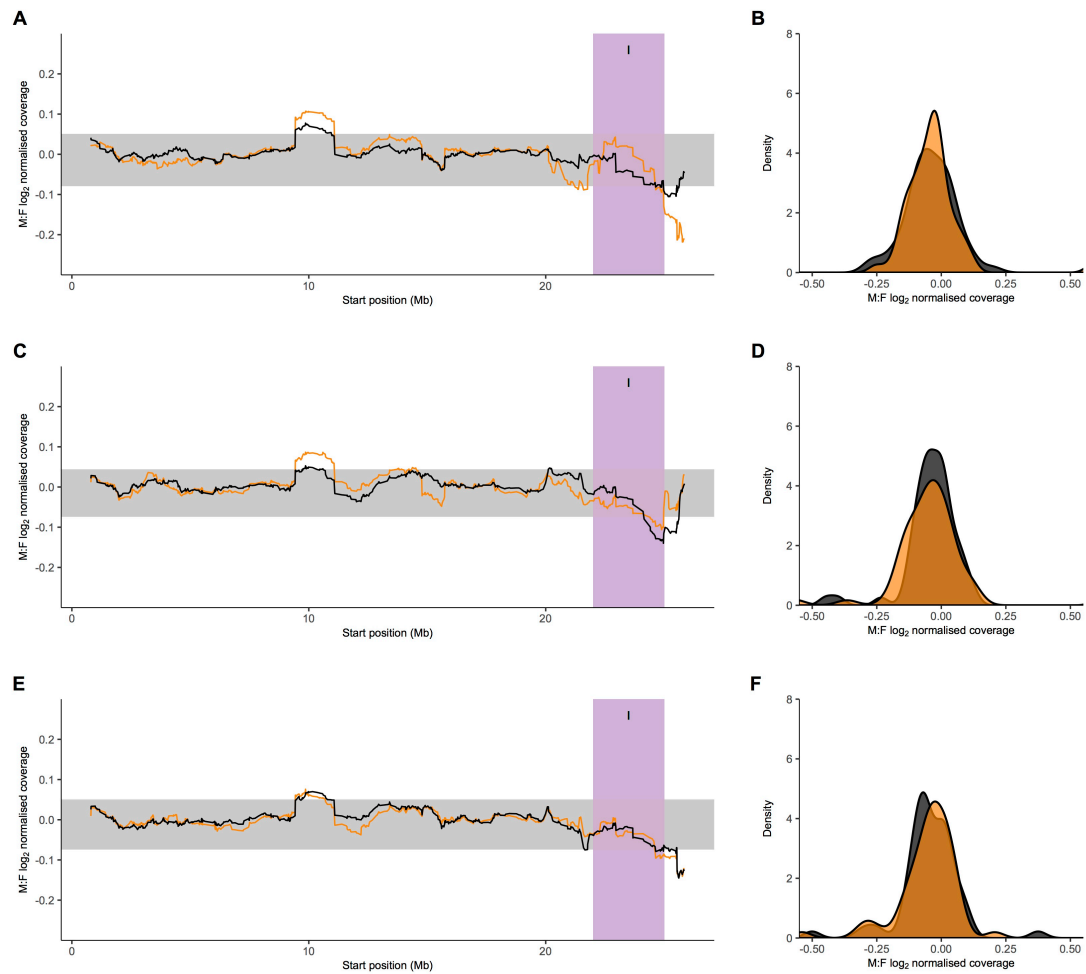
Supplementary Fig. 7. Male:female coverage for the X chromosome across upstream (orange) and downstream (black) guppy populations. Panels A. C. and E. Moving averages of normalised coverage across the X chromosome based on sliding window analysis (window size of 40 scaffolds) for Yarra (panel A), Quare (panel C) and Aripo (panel E) watersheds. 95% confidence intervals based on bootstrapping autosomal estimates are in grey. Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in laboratory population males (Stratum I, 22-25 Mb) (see Fig. 2), light purple indicates the region with less X-Y differentiation, where there a significant excess of male SNPs in laboratory populations (Stratum II, 15 – 22 Mb). Panels B. D. and F. Distribution of sex differences in normalised coverage for the oldest region of the X chromosome (Stratum I, 22 – 25 Mb) for Yarra (panel B), Quare (panel D) and Aripo (panel F) watershed.