Evolution of the canonical sex chromosomes of the guppy and its relatives

M. Kirkpatrick[1,*,†], G. Dixon[1], J.M. Sardell[1], M. Schartl[2], C. L. Peichel[3]

[1] Dept. of Integrative Biology, University of Texas, Austin TX 78712 USA

[2] Developmental Biochemistry, University of Würzburg, Biozentrum, Am Hubland, 97074 Würzburg, Germany and The Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas, TX 78666, USA

[3] Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern Switzerland

* Corresponding author: kirkp@mail.utexas.edu, +1-512-499-0825

[†] Authors listed in order of their relative contributions.

Abstract

The sex chromosomes of the guppy, *Poecilia reticulata*, and its close relatives are of particular interest: they are much younger than the highly degenerate sex chromosomes of model systems such as mammals and *Drosophila melanogaster*, and they carry many of the genes responsible for the males' dramatic coloration. Over the last decade, several studies have analyzed these sex chromosomes using a variety of approaches including sequencing genomes and transcriptomes, cytology, and linkage mapping. Conflicting conclusions have emerged, in particular concerning the history of the sex chromosomes and the evolution of suppressed recombination between the X and Y. Here we address these controversies by reviewing the evidence and reanalyzing data. We find no support for a nonrecombining sex determining region (SDR) or evolutionary strata in *P. reticulata.* We confirm that its congener *P. picta* has evolved dosage compensation across all of its X chromosome. Last, we do not find evidence that the nonrecombining SDRs of *P. picta* and *P. wingei* descend from a common ancestral SDR, and suggest instead that suppressed recombination between the X and Y evolved independently after the two species diverged. We identify possible causes of conflicting results in previous studies and suggest best practices going forward.

1: INTRODUCTION

The origin and evolution of young sex chromosomes are of particular interest to evolutionary genomics. They are the most rapidly evolving part of the genome in many animals and plants, and they have evolutionary features that give unique insights into the evolution of recombination, sexually antagonistic selection, and other important processes (Bachtrog *et al.* 2011). The guppy, *Poecilia reticulata*, holds a special place in the history of this subject. Because they carry most of the genes responsible for the males' famed coloration, the *P. reticulata* sex chromosomes have been studied since the 1920s (Winge 1922; Haskins *et al.* 1961; Lindholm & Breden 2002; Charlesworth 2018). The last decade has seen a burst of research on the sex chromosomes of *P. reticulata* and its relatives. Several recent studies have arrived at conflicting conclusions, notably regarding the evolution of recombination. These controversies are the focus of this paper.

Studies from the pre-genomic era provided conflicting conclusions regarding recombination between the sex chromosomes. Using linkage maps, Tripathi *et al.* (2009) reported that recombination between the X and Y of *P. reticulata* is confined to a relatively small region bounded on either side by large nonrecombining regions. Nanda *et al*. (2014) used cytology and linkage maps to study the sex chromosomes of *P. reticulata*, its sister species *P. wingei*, and the closely related *P. obscura*. They concluded that the Y chromosomes of these three species descended from a common ancestral Y, and that the X and Y chromosomes of *P. reticulata* recombine down their lengths (save perhaps a small heterochromatic region specific to the Y). Further, they reported that an extended region of heterochromatin at the end of the Y opposite to the centromere has evolved in *P. wingei* that blocks recombination with the X. Again using cytology, Lisachov *et al.* (2015) found that recombination between the X and Y in *P. reticulata* is rare and concentrated towards the end of the chromosome opposite to the centromere.

The pace of discovery accelerated with the arrival of genome sequences for *P. reticulata* (Fraser *et al.* 2015; Künstner *et al.* 2016). By analyzing molecular variation among resequenced genomes from several natural populations of *P. reticulata*, Wright *et al.* (2017) drew conclusions at odds with the previous reports: crossing over between the X and Y is completely

blocked over about 40% (10 Mb) of the chromosomes, with recombining regions on either side. They reported that the nonrecombining sex determining region, or SDR, is divided into two strata, which are regions in which crossing over between the X and Y was completely suppressed at different times in the past (Lahn & Page 1999; Charlesworth 2017). Wright *et al.* (2017) further concluded that the nonrecombining SDR has expanded independently in three populations that inhabit the headwaters of rainforest streams. Morris et al. (2018) followed up this study by identifying 40 loci that are unique to the nonrecombining region of the Y chromosome, and proposed two of them as candidates for the sex determining gene.

In the next study from the same research team, Darolti *et al.* (2019) enlarged the phylogenetic picture by analyzing genomes and transcriptomes from five species of poeciliid fish, including *P. reticulata*, its sister species *P. wingei*, and the congener *P. picta*. These authors found that Chromosome 12 is responsible for sex determination in all three species. They reported that the two strata on the *P. reticulata* Y are shared with *P. wingei*, and so they evolved in the common ancestor of those species. In the more distantly related *P. picta*, they found the Y chromosome to be highly degenerate, and provided evidence that the X chromosome has evolved chromosome-wide dosage compensation of gene expression in response. They also concluded that the SDRs in all three species descend from a common ancestor, which implies that the rates at which the Y degenerates varies greatly between lineages. Darolti *et al.* (2020) concluded from linkage mapping that completely suppressed recombination is confined only to the first of the two strata in *P. reticulata*, while there is very rare crossing over between the X and Y in the second "stratum". In the most recent paper from that research group, Almeida *et al.* (2020) carried out long-read sequencing on much larger samples of individuals from six natural populations. They also concluded that there is no recombination in Stratum 1.

An independent research team studied the *P. reticulata* sex chromosomes using linkage mapping (Bergero *et al.* 2019). They rejected the hypothesis that crossing over between the X and Y is completely suppressed anywhere on these chromosomes. They reported that males have extremely low recombination rates on all chromosomes (autosomes as well as the sex chromosomes) except near the telomeres. Recombination between the X and Y in natural

populations is apparently very rare since there is elevated F_{ST} between males and females down the entire length of Chromosome 12, in some regions attaining values greater than 0.1. This study also identified a region of high recombination on the end of the Y chromosome opposite to the centromere, consistent with the findings of Lisachov *et al.* (2015). The localization of crossovers in males to the tips of all chromosomes is consistent with the high GC content found there (Charlesworth *et al.* 2020b). Charlesworth *et al.* (2020a) reported a crossover between the X and Y of *P. reticulata* in a region previously hypothesized to be a nonrecombining stratum. Most recently, Fraser et al. (2020) assembled a new high quality reference genome for *P. reticulata* that includes both the X and Y, and they resequenced fish from six natural populations. They concurred with Bergero *et al.* (2019) that there is no evidence for large nonrecombining strata. Further, they found two candidate regions for the sex determining gene that (surprisingly) are located at opposite ends of the Y chromosome. Figure 1 summarizes some of these results.

These conflicting conclusions have led to controversy and confusion in the scientific community (Wright *et al.* 2019; Bergero & Charlesworth 2019). In an effort to remedy this situation, in this paper we review, reanalyze, and reinterpret published data from *P. reticulata* and its close relatives, *P. wingei* and *P. picta*. We focus on the evolution of recombination, the evolution of dosage compensation, and the origin of the Y chromosomes. We find no evidence for strata in *P. reticulata* or for a region where crossing over between the X and Y is completely suppressed. Our results confirm that much of the X and Y in *P. wingei* and *P. picta* have extremely low or no recombination, and that the Y in *P. picta* is highly degenerate. The X chromosome of *P. picta* does indeed show strong evidence of dosage compensation.

Last, we consider the hypotheses regarding homology of the Y chromosomes and SDRs in all three species. While the homology of the Y chromosomes of *P. reticulata* and *P. wingei* has been established by molecular cytogenetics (Nanda *et al.* 2014), no genetic evidence has been adduced regarding homology of the Y in *P. picta*. An alternative hypothesis to homology is that there has been a turnover event in which an ancestral Y was replaced by a neo-Y derived from an X chromosome in the ancestor of *P. wingei* and *P. reticulata* after divergence from *P. picta* (van Doorn & Kirkpatrick 2007; Bergero & Charlesworth 2019; Vicoso 2019; Meisel 2020). This

could explain why the Y of *P. picta* Y chromosome is so much more degenerate than the Ys in the other two species. Gene trees do not support either the homologous SDR or turnover hypotheses. We conclude that the SDRs most likely evolved independently in *P. picta* and the *reticulata-wingei* lineage after they diverged.

2: METHODS

DNA and RNA sequences were obtained by Darolti *et al.* (2019) from three males and three females of *P. picta* sampled from nature, and from three males and three females of *P. wingei* from a strain acquired from a fish fancier. The data for *P. reticulata* were obtained by Wright *et al.* (2017). DNA sequences are of two males and two females from a lab population and 24 males sampled from nature. RNA sequences are of 11 males and four females from the same lab population.

We downloaded sequencing reads from Bioprojects PRJNA528814 and PRJNA353986 on the SRA database. Multiple fastq files for the same individual were concatenated based on Sample Name in the SRA metadata. Read quality was assessed using FastQC (Andrews 2010). Reads were trimmed for adapter sequences and low quality bases using cutadapt (Martin 2011).

We chose to map DNA and RNA reads against the *Xiphophorus maculatus* reference genome (Version 5, Schartl *et al.* (2013)) for two reasons: it is by far the highest quality genome among poeciliid fishes, and this species is an equal phylogenetic distance from all three *Poecilia* species that are our focus. We acquired the sequence from Ensembl (ftp://ftp.ensembl.org/pub/release-99/fasta/xiphophorus_maculatus/dna/) and mapped DNA and RNA reads to the reference with Bowtie2 using default parameters and the –local argument (Langmead & Salzberg 2012). PCR duplicates were removed using Picard Toolkit (Broad Institute 2019). Alignment files were sorted and subsetted by chromosome using SAMtools (Li *et al.* 2009).

There is some uncertainty about the physical coordinates on the *P. reticulata* sex chromosomes. Reference genomes from three different species were used to map sequencing reads by the Mank team (e.g. Wright *et al.* (2017)), the Charlesworth team (Bergero *et al.* 2019), and this study (see the Discussion). Further, there are several reports of inversions and assembly errors on the *P. reticulata* sex chromosomes (Nanda *et al.* 2014; Bergero *et al.* 2019; Darolti *et al.* 2020; Charlesworth *et al.* 2020a; Fraser *et al.* 2020). Accordingly, we view the coordinates as including error and interpret them with caution.

Fold coverage from DNA reads was averaged in 10 kb windows using BEDtools (Quinlan & Hall 2010). Genotyping and quality filtering (QUAL < 20) was performed using mpileup and BCFtools (Li 2011). Genotypes for DNA and RNA alignments from all species were called together. We used VCFtools to remove indels, singletons, and select for biallelic SNPs. We required a minimum mean depth of 3. To assess gene expression, fold coverage for RNA reads mapping within gene boundaries were counted using FeatureCounts (Liao *et al.* 2014).

To determine which alleles are X- and Y-linked, we considered patterns of heterozygosity in the genomic (DNA) sequences. An allele was considered to be putatively Y-linked if it always appears in males in heterozygotes and was absent from females (which were homozygous at these sites since we used only bi-allelic SNPs). We designate these alleles as “Y-like”, the alternate alleles as “X-like”, and the sites at which these occur as “SDR-like SNPs”. We imposed the additional criterion that these sites have data from at least two individuals of each sex. Because the sample sizes are small, SDR-like SNPs can occur by chance at autosomal loci. For that reason, we expect that the loci identified as sex linked by this analysis may include false positives. Nevertheless this set of loci will be highly enriched for genes that truly are sex-linked.

To prepare genotypes for calling X- and Y-linkage, we first split the genotypes by species using VCFtools. We then filtered SNPs to include only those with no missing individuals (--max-missing 1.0) and a minimum allele frequency of 0.2. We then split the genotypes further by sex, and calculated allele frequencies using --freq in VCFtools and observed heterozygosities using --hardy in VCFtools. The rules outlined above for identifying the sex linkage of alleles were implemented in a custom R script. We applied this analysis to the autosomes as well as the sex chromosomes in order to assess how frequently genes with apparent sex linkage occur

throughout the genome. Once the putative X- and Y-linked alleles were identified, their depths from both the DNA and RNA reads were isolated from the VCFs using a custom Python script.

We analyzed dosage compensation in *P. picta* using two strategies. First, we determined the ratio of read depths in males and female within coding regions for both genomic DNA and RNA transcripts. We then compared these ratios for all loci on autosomes and all loci on the sex chromosomes. Second, we identified the SDR-like SNPs, then calculated the ratio of read depths between the X-like and Y-like alleles for both genomic DNA and RNA transcripts. Results were calculated separately for genes in the SDR and PAR, and we placed the boundary between these two regions at 20.7 Mb based on visual inspection of the statistics shown in Figure 2. The SDR-PAR boundary in *P. wingei* appears to be in approximately the same region, so we partitioned its read depth data into two regions as for *P. picta*.

We used gene trees to test hypotheses regarding the evolution of suppressed recombination. Unfortunately, the extreme degeneration of the Y chromosome in *P. picta* greatly complicates the analysis. At hemizygous sites (where there is a deletion on the Y), BCFtools (Li 2011) and other widely-used SNP calling software will impute a homozygote and assign the X-linked allele to the Y. This error results in gene trees in which the Xs and Ys from *P. picta* appear to cluster together to the exclusion of the Xs and Ys of *P. wingei*, which could erroneously suggest that the SDRs evolved independently or that there was a turnover in *P. picta*. To avoid this bias, we focused on a subset of the SDR-like SNPs in *P. picta* that are free from this artifact. Alleles were polarized as either ancestral or derived using *P. latipinna* and *Gambusia holbrooki* as outgroups (data from Darolti *et al.* (2019)), and we filtered out sites at which those species are not fixed for the same allele. We included only sites where all *P. wingei* males have the same genotype (heterozygote or homozygote), and all *P. wingei* females have the same homozygote genotype. Finally, we discarded sites at which the derived allele is unique to *P. picta* because they are not informative regarding the homology of the Y chromosomes. We refer to the SNPs that meet all of these criteria as "topologically informative." Unfortunately, these criteria exclude gene trees that would result if the SDRs in the two species had independent origins, that is, trees in which the Xs and Ys from *P. picta* cluster together, as do the Xs and Ys of *P. wingei.*

3: RESULTS

3.1: Divergence and recombination between the X and Y

To study divergence between the X and Y chromosomes, we computed six statistics in sliding windows: the ratio of read depth in males and females ("read depth ratio"), the ratio of the density of all SNPs in males and in females ("SNP density ratio"), the density of SNPs with patterns of heterozygosity consistent with the SDR ("SDR-like SNP density"), the density of SNPs with female-specific alleles ("female-specific SNP density"), F_{ST} between males and females using Weir and Cockerham's (1984) estimator ("F_{ST}"), and the ratio of gene expression in males and females ("expression ratio"). The windows were 10 kb for the read depth ratio and 100 kb for the other five statistics. Figure 2 shows the results for all six statistics from the three species.

In *P. reticulata*, there is no sign that the X and Y have genetically diverged anywhere along their lengths. We do not see a decreased read depth ratio in the genomic data (a telltale signature of Y chromosome degeneration), and F_{ST} between males and females is close to zero everywhere. The patterns for these statistics are very similar using the RNA sequences, which include a larger sample of females. The other four statistics fall within the ranges typical of autosomes. These results do not provide any evidence of a nonrecombining sex determining region (SDR).

Our results are consistent with the results but not the interpretation of previous analyses that used a second signature of X-Y recombination (Almeida *et al.* 2020; Darolti *et al.* 2020). Following the origin of a non-recombining stratum by an inversion (or any other mechanism involving a *cis* recombination modifier), all of the Y chromosomes that inherit that stratum will form a monophyletic clade with respect to the X chromosomes (Dixon *et al.* 2018; Toups *et al.* 2019). The monophyly persists through speciation events: the homologous strata on the Ys from all the descendant species continue to form a clade with respect to the X chromosomes, and the monophyly also extends across the entire length of the stratum. In recombining

regions of the sex chromosomes, however, this monophyly breaks down. Recombination causes gene copies from the X and Y chromosomes of each species to be intermingled on a gene tree. Recombination events more recent than a speciation event cause gene copies from the X and Y of one species to cluster together rather than with their gametologs in other species. Gene trees thus offer a sensitive way to distinguish the PAR from the SDR because they integrate signals of recombination that have accumulated in natural populations over many generations (Dixon *et al.* 2018; Toups *et al.* 2019).

Darolti *et al.* (2020) estimated the gene trees at 42 loci spread along the length of the sex chromosomes. Figure 1 shows their locations. One of these loci, *alad*, shows a gene tree with the monophyletic Y topology consistent with a nonrecombinating stratum shared between *P. reticulata* and *P. wingei*, but the statistical support for its monophyly is weak (bootstrap value: 51%). At the locus *npr2*, which is less than 3 kb away from *alad*, there is strong statistical support for the node joining the X and Y in *P. reticulata* together (bootstrap value: 97%) and for a node joining the X and Y in *P. wingei* (bootstrap value: 96%). This is evidence that in both species there has been recombination between the X and Y chromosomes in the region between the *npr2* locus and the sex determining region since the two species diverged. The remaining 40 gene trees are also inconsistent with a stratum that predates the divergence of *P. reticulata* and *P. wingei*. Most recently, Almeida *et al.* (2020) used a method developed by Dixon *et al.* (2018) to analyze gene trees with much larger samples. In five of the six populations studied, the majority of trees in the proposed Stratum 1 and elsewhere on Chromosome 12 have topologies that are consistent with ongoing recombination between the X and Y in *P. reticulata*.

In the sister species, *P. wingei*, we find that the picture is quite different. Three of the statistics (SNP density ratio, SDR-like SNP density, and F_{ST}) fall far outside the autosomal range of values over much of the center part of the sex chromosomes (Figure 2). These patterns are consistent with very little or no recombination. The read depth ratio remains near 1, however, suggesting that there has not been extensive degeneration on the Y chromosome. These conclusions concur with Darolti *et al.* (2019).

Within the region of reduced recombination, there is a smaller segment between about 10 Mb and 20 Mb where the female-specific allele density is greatly elevated, with very sharp boundaries at each end. The other three statistics just mentioned show no unusual patterns specific to this 10 Mb region. We speculate that these patterns might result from a polymorphic inversion on the X chromosome of *P. wingei* that is evident in the cytological data of Nanda *et al.* (2014).

The third species, *P. picta*, shows yet another distinctive set of patterns. Most strikingly, the ratio of read depths for genomic DNA in males and females is about one half over most of the proximal end of the sex chromosome (Figure 2). This is indicative of large-scale deletions and/or divergence of the Y sequence to the point that reads from it no longer map to the reference. The SDR-like SNP density, female-specific allele density, and F_{ST} are much lower than in *P. wingei*, presumably for the same reasons. These conclusions again agree with Darolti *et al.* (2019).

3.2: Dosage compensation

Figure 3 shows that for genes on the sex chromosomes of *P. wingei*, the ratio of read depths in males and females for genomic DNA and RNA transcripts are near to 1 in coding regions on both autosomes and the sex chromosomes. Results for *P. reticulata* are very similar to those shown for *P. wingei*. This suggests there has been little or no degeneration affecting gene expression on the Y of those two species. In *P. picta*, by contrast, the read depth for genomic DNA in the SDR of *P. picta* males is half of its value in females (as noted earlier). Read depth for the DNA in the PAR is also reduced in males (but much less so than in the SDR), which could result because the region we identified as PAR includes some of the SDR. Nevertheless, the total gene expression in males and females is similar in both the SDR and PAR (Figure 3). This strongly suggests that there is dosage compensation in which most or all genes on the X chromosome have doubled their expression to compensate for loss of expression from the Y, supporting the conclusions of Darolti *et al.* (2019).

A complimentary perspective comes from the ratio of read depths for alleles we impute to be X-linked or Y-linked (see Methods). In *P. reticulata* and *P. wingei*, the Y:X read depth ratio is near 1 for both genomic DNA and RNA transcripts (Figure 4). That is consistent with the earlier conclusion that there is little or no degeneration of Y-linked genes in those species. In *P. picta*, however, the distribution of that ratio for genomic DNA is skewed. There is a peak near to 1, indicative of coding loci that are present on both the X and Y (many likely to be in the PAR). But there are many loci where the Y:X ratio is substantially smaller than 1. This could result if Y-linked reads are highly divergent from the X sequence such that very few of them map to the reference genome. The RNA transcripts show there are many genes at which the Y-linked allele has very little or no expression, again consistent with substantial degradation of the Y chromosome, and with the evolution of a dosage compensation mechanism because the overall expression of these genes does not differ between males and females (Figure 3).

3.3: Evolution of suppressed recombination on the Y chromosomes

We used gene trees to investigate the evolution of suppressed recombination on the Y chromosomes. We focused on *P. wingei* and *P. picta* because we found no evidence of a nonrecombining SDR in *P. reticulata*. One hypothesis is that the oldest parts of the SDRs on the Y chromosomes of *P. wingei* and *P. picta* are homologous, meaning that the SDRs descend from an ancestral SDR on a Y chromosome in their common ancestor. This possibility is favored by Darolti *et al.* (2019). An alternative hypothesis is that there has been a turnover in which the Y chromosome of *P. wingei* was derived from an X, an idea favored by Bergero and Charlesworth (2019) and Meisel (2020). A third hypothesis is that the sex chromosomes of the ancestor of both species were largely recombining and that their SDRs evolved independently when recombination was suppressed after the species diverged.

These hypotheses make contrasting predictions regarding the gene trees from the SDRs (Figure 5). If the Y chromosomes of *P. wingei* and *P. picta* descend from a single common ancestral Y, then in the gene trees from the SDRs of the sex chromosomes the Y chromosomes from both species will cluster together to the exclusion of the X chromosomes (Tree 1 in Figure 5) (Dixon *et al.* 2018). If the Y of *P. wingei* was derived from its X chromosome in a turnover

event after that lineage diverged from *P. picta*, then the X and Y chromosomes of *P. wingei* will cluster together and form a sister clade to the *P. picta* X chromosomes (Figure 2) (Sardell *et al.* 2020). If the SDRs evolved independently in the two species, then the X and Y of each species will cluster together in gene trees from their SDRs. Other evolutionary histories (*e.g.* a turnover in *P. picta* or introgression of sex chromosomes between species) result in yet other topologies.

As explained in the Methods, the degeneration of the Y chromosome in *P. picta* can cause hemizygous sites to be erroneously interpreted as homozygous and lend false support to the independent origins and turnover hypotheses. To avoid this bias, we focused on SNPs that are free from this artifact and which we refer to as "topologically informative" (see Methods). While some of these SNPs are found on autosomes, they are much abundant on the sex chromosomes (Figure 6). There are 208 SNPs on the sex chromosomes of *P. picta* that are topologically informative, and they are spread quite evenly over the chromosome. We used these SNPs to construct gene trees.

Only two of the 208 gene trees show the topology expected if the SDRs of *P. picta* and *P. wingei* are homologous (Figure 5). At 10 SNPs, the gene tree is consistent with a turnover that occurred in the ancestor of *P. wingei*. But at 194 SNPs, the X and Y of *P. wingei* both share a derived allele with the Y of *P. picta*. This configuration could result from a turnover in which the *P. wingei* X was derived from a Y, which is very unlikely if the ancestral Y was degenerate. It could also result from introgression of the Y from the ancestor of *P. wingei* into *P. picta*, or from introgression of the X from an outgroup species into *P. picta*. Although introgression of sex chromosomes between species has been observed in stickleback fishes (Dixon *et al.* 2018), it is thought to be extremely rare. Further explanations for Tree 3 are incomplete lineage sorting and sequencing error, but these should result in equal frequencies of the Tree 2 and Tree 3 topologies. A final hypothesis for Tree 3 topology is homoplasy with much higher mutation rates on the Y than on the X in *P. picta*. The last topology is Tree 4, which is seen at two SNPs. This outcome has no simple historical interpretation and likely results from genotyping errors or homoplasy.

Overall, the data do not support the hypothesis that the SDRs in *P. picta* and *P. wingei* descend from a common ancestral SDR, nor do they support the hypothesis of a turnover in

which a new Y evolved from the X in the ancestor of *P. wingei* and *P. reticulta*. It is very plausible that their SDRs originated independently by suppression of recombination after the two lineages split. Unfortunately, gene trees cannot provide unambiguous support for this hypothesis because of the errors caused by hemizygous sites described above. Several hypotheses might explain the stark contrast in the extent of Y degeneration between the two species. In descending order of plausibility, the data are consistent with differences in when recombination was suppressed in the two lineages, differences in the rates of degeneration of their Ys after recombination suppression, or the independent recruitment of Chromosome 12 as a sex chromosome from an autosome.

4: DISCUSSION

4.1: Current state of the X and Y chromosomes in the three species

The results presented above and those from previous studies provide several insights into the state of the sex chromosomes in these three species. In the guppy *P. reticulata*, four lines of evidence argue that the sex chromosomes have not evolved suppressed recombination or evolutionary strata and in fact continue to recombine in the region near to the sex determining gene. First, we see no signs of increased divergence between males and females (a proxy for divergence between the X and Y) in the candidate regions for sex determination or anywhere else along the chromosome (Figure 2). Second, gene trees on the sex chromosomes are consistent with crossovers that occurred in the region identified as nonrecombining by Wright *et al.* (2017) and later studies by that research team (Morris *et al.* 2018; Darolti *et al.* 2019; Darolti *et al.* 2020; Almeida *et al.* 2020). We expect that the most recent recombinant Y chromosome was established within the last few thousand generations based on the effective population size of Y chromosomes relative to autosomes, and the small census population sizes of guppies (Fraser *et al.* 2015). Third, a crossover has been directly observed in one of the regions that had previously been proposed as a nonrecombining stratum (Bergero *et al.* 2019;

Charlesworth *et al.* 2020a) (see Figure 1). These latter two studies also found that very rare recombination except near the ends of chromosomes in males is not a unique feature of the sex chromosomes, but is found on autosomes as well, consistent with earlier reports in guppies (Tripathi *et al.* 2009; Lisachov *et al.* 2015). Fourth, SNPs with heterozygosity patterns consistent with sex linkage are scattered across the length of the sex chromosomes (Bergero *et al.* 2019), rather than being concentrated near the sex determining factor or restricted to the proposed nonrecombining region.

Apparently recombination in males is sufficiently rare in at least some populations that the X and Y chromosomes have diverged at the molecular level, *e.g.* F_{ST} between males and females in the range 0.05 to 0.1 over large regions of the sex chromosomes (Bergero *et al.* 2019; Almeida *et al.* 2020). For unknown reasons, our analyses do not show comparable differentiation (Figure 2). We concur with Bergero *et al.* (2019) that nothing suggests that patterns of recombination on the sex chromosomes of *P. reticulata* are different than those on the autosomes.

While the degree of sex bias in recombination (heterochiasmy) is extreme in guppies, qualitatively similar differences between males and females are seen across the eukaryotes (Sardell & Kirkpatrick 2020). In fact, there are examples of heterochiasmy even more extreme than guppies. In hylid frogs, recombination in males continues along almost the entire lengths of all chromosomes (Brelsford *et al.* 2016; Rodrigues *et al.* 2017), but it is about 10^5 less frequent in males than in females (Guerrero *et al.* 2012). Recombination also occurs in male ranid frogs but is so rare that in some populations that F_{ST} between males and females approaches its maximum possible value (Jeffries *et al.* 2018; Toups *et al.* 2019).

Variation between populations in recombination rates on the sex chromosomes is a recurring theme in the guppy literature. Using laboratory crosses, Haskins *et al.* (1961) showed a significant difference between two populations in the recombination rate between the sex-linked locus *Sb* and the sex determining gene. Fraser *et al.* (2015) reported differences between populations in the linkage of certain color patterns to the X and Y chromosome, but their findings regard differences in the frequencies of color pattern alleles on the X and Y chromosomes rather than differences in recombination rates. In contrast, Wright *et al.* (2017)

and Almeida *et al.* (2020) argue that differences between populations in patterns of molecular variation reflect differences in the actual rates of recombination between the X and Y.

We suggest that subtle differences between populations in the degree of molecular differentiation between the guppy X and Y could occur even in the absence of differences in the recombination landscape. Consider the consequences of a rare crossover between the X and Y, for example near to the male determining factor. If the new Y haplotype spreads to high frequency by selection or drift, differentiation between the X and Y will be erased from the crossover breakpoint down the rest of the chromosome distal to the male determining gene. In the headwaters of the streams where they live, population sizes of *P. reticulata* are only a couple thousand individuals (Fraser *et al.* 2015). Consequently, crossovers between the X and Y may occur very infrequently, giving the X and Y time to diverge slightly before the next recombinant Y chromosome is established. Downstream populations are many times larger (Fraser *et al.* 2015), so recombinant Y chromosomes appear much more frequently and so divergence between the X and Y has less opportunity to develop.

A second factor that could contribute to patterns of molecular variation is sexually antagonistic selection, or SAS. More than 50 traits that are under sexual selection in male *P. reticulata* have been mapped to the sex chromosomes (Lindholm & Breden 2002). These are expected to generate peaks in the divergence between the X and Y in neutral genetic variation (Kirkpatrick & Guerrero 2014) and can generate patterns that give the appearance of suppressed recombination (Charlesworth 2018; Bergero *et al.* 2019). Recombination in *P. reticulata* males is so rare that the map length of almost the entire Y chromosome is much less than 1 cM (Bergero *et al.* 2019). The small census population sizes suggest that many targets of SAS may lie within less than one ρ (= $N_e r/2$) of the sex determining factor, a condition favorable to inflating F_{ST} between the X and Y (Kirkpatrick & Guerrero 2014). Indeed, peaks in F_{ST} between males and females (a proxy for divergence between the X and Y) consistent with SAS have been reported by Bergero *et al.* (2019) and are visible in Figure 1 of Almeida *et al.* (2020). Thus differences between populations in the intensity of sexual selection and the frequencies of alleles at loci that experience SAS can also contribute to the differences between populations in the degree of X-Y divergence even in the absence of variation in recombination rates.

The very recent publication of a greatly improved reference genome for *P. reticulata* that includes the Y (Fraser *et al.* 2020) adds additional insight. That paper concludes that if an SDR exists, it is very small (less than 1 Mb). The most likely candidate for the sex determining region is absent from the X chromosome, which would explain why it was not detected previously in studies based on earlier reference genomes that lacked the Y. It may be possible to distinguish which of the two candidate regions identified by Fraser *et al.* (2020) determines sex using the gene tree method developed by Dixon *et al.* (2018).

In the guppy's sister species, *P. wingei*, molecular divergence between males and females suggests that the X and Y have extremely low or no recombination over most of their lengths. Divergence is clearly evident in four of the six statistics shown in Figure 2. Within that large region, there is a smaller segment of about 10 Mb that shows greatly elevated female-specific allele density. This region could reflect a polymorphic inversion on the X chromosome of *P. wingei* (Nanda *et al.* 2014). If so, the patterns seen in Figure 2 could result if the three X chromosomes in our sample of *P. wingei* males are monomorphic for one arrangement while the six X chromosomes in the females include both arrangements. We are not able to determine if there is more than one stratum blocking recombination between the X and Y in *P. wingei*. A segment of the sex chromosomes opposite to the centromere shows no sign of molecular divergence between the X and Y, suggesting high rates of recombination there.

The most dramatic patterns are seen in *P. picta*. Consistent with the findings of Darolti *et al.* (2019), our analyses suggest that most of its Y chromosome no longer recombines, and much of it has degenerated by sequence divergence or deletion (Figure 2). We also support their conclusion that genes across the X chromosome have evolved dosage compensation, that is, higher expression to compensate for the degeneration of the Y.

4.2: History of the sex chromosomes

We find no support for the hypothesis that X-Y recombination was repressed long ago over some or much of the ancestral sex chromosomes of *P. wingei, P. picta*, and *P. reticulata.* Nor do we find support for the hypothesis that a turnover occurred in the common ancestor of *P. reticulata* and *P. wingei* after that lineage diverged from *P. picta*. Although direct evidence is

lacking, we feel that the most likely hypothesis is that suppressed recombination evolved independently in *P. picta* and *P. wingei* after those lineages diverged. The Y may be much more degenerate in *P. picta* than in *P. wingei* simply because recombination between its X and Y was suppressed much earlier. If the Y originally evolved in their common ancestor, that would imply extreme variation in the evolution of suppressed recombination over the same time scale. But because we find no evidence of a shared SDR, we are not able to rule out the possibility that the same autosome independently evolved into a sex chromosome in *P. picta* and the ancestor of the other two species. While this hypothesis seems unparsimonious, the independent recruitment of the same autosome as a sex chromosome has been observed in several animal taxa (Jeffries *et al.* 2018; Dixon *et al.* 2018; Sardell *et al.* 2020).

4.3: Reconciliation of past studies and best practices going forward

How can the many discrepancies between the conclusions from previous studies be reconciled, and how best can the field move forward? Some of the problems can be traced to semantic differences. We defer to the terminology for sex chromosomes defined by Bachtrog *et al.* (2011). The sex determining region, or SDR, is a segment of the sex chromosomes that includes the sex determining factor and in which recombination between the X and Y is completely suppressed. Within the SDR, there can be one or more strata, which are regions in which crossing over was entirely suppressed at different times in the past (Lahn & Page 1999; Charlesworth 2018). The SDR may be as small as a nucleotide, as in the case of the fugu, or as large as most of the Y chromosome, as in mammals (Bachtrog *et al.* 2014). By this definition, it is a *non sequitur* to say there is a stratum on the guppy sex chromosomes in which there is rare recombination (Darolti *et al.* 2020). Again following Bachtrog *et al.* (2011), all of the sex chromosomes that fall outside of the SDR make up the pseudoautosomal region, or PAR, regardless of whether the local recombination rate (measured as cM/Mb) is very high, as in mammals, or very low, as in some frogs (Bachtrog *et al.* 2014). By this definition, it is also a *non sequitur* to refer to only part of the recombining region as the PAR (Bergero *et al.* 2019). Authors are of course free to adopt noncanonical definitions, but in that case we urge that they give explicit definitions and use terms consistently.

There are, however, at least three substantive reasons why different studies have arrived at differing conclusions. First, some statistical approaches used to define the SDRs and PARs of the guppy and its relatives are problematic. Two nonrecombining strata were identified in *P. reticulata* and *P. wingei* using the criterion that molecular differences between males and females over a region of the sex chromosome fall outside the 95% interval seen on autosomes (Wright *et al.* 2017; Darolti *et al.* 2019). By that standard, 5% of the entire genome is expected to be identified as nonrecombining sex chromosome strata, and indeed several regions of autosomes do meet that criterion (Supplementary Figure 1 in Wright *et al.* 2017). Further, the differences between males and females used to define nonrecombining strata in those studies are extremely small and likely not biologically meaningful (*e.g.* a male:female SNP density ratio differing much less than 1% from the autosomal average (Wright *et al.* 2017)). Finally, three mapping studies have given unambiguous evidence that the Y chromosome of *P. reticulata* recombines in the region that had been proposed as a nonrecombining stratum (Figure 1) (Bergero & Charlesworth 2019; Charlesworth *et al.* 2020a; Darolti *et al.* 2020). Thus the statistical strategy used by Wright *et al.* (2017) and Darolti *et al.* (2019) is inadequate to define SDRs and strata.

Second, the results from mapping experiments have been interpreted in different ways. It is difficult to draw conclusions from a failure to observe crossovers, especially in crosses with few offspring. Mapping experiments are underpowered to distinguish between partial and full linkage of loci to the male determining gene (Muyle *et al.* 2016; Wright *et al.* 2019). This limitation is particularly acute in species with extremely low recombination rates in males, such as the guppy. Conclusions about regions where the X and Y do not recombine based on mapping (Darolti *et al.* 2020) should be interpreted with caution. Methods that define nonrecombining sex determining regions using gene trees are much more sensitive because they integrate genetic signatures of recombination over long periods of evolutionary time (Dixon *et al.* 2018; Toups *et al.* 2019).

Third, sequencing studies have used different strategies to map reads. Wright *et al.* (2017) used a *de novo* assembly of *P. reticulata* whose scaffolds were then ordered according to the guppy reference (N50 = 0.017 Mb). Bergero *et al.* (2019) and Charlesworth *et al.* (2020a)

mapped their DNA reads to the female guppy reference genome (with N50 = 5.3 Mb; Künstner *et al.* (2016)). While this reference has the obvious advantage of being one of the species of interest, it is based entirely on short-read sequencing. Darolti *et al.* (2019; 2020) used one of the publicly-available *Xiphophorus helleri* genomes (with N50 = 29.4 Mb; Shen *et al.* (2016)). Last, for this paper we used the *Xiphophorus maculatus* reference because it has the highest quality of any species in this family (Schartl *et al.* 2013). The current assembly (version 5) is based on long-read as well as short-read sequencing and BioNano chromosome assembly, and has an N50 scaffold length of 31.5 Mb. We note that of these reference genomes, only the *X. maculatus* reference that we used includes the Y chromosome. Mapping artifacts that occur when the reference genome is missing the Y lead to bioinformatic errors, such as inflated estimates for F_{ST} between males and females, on autosomes as well as the sex chromosomes (Bissegger *et al.* 2019; Cheng & Kirkpatrick 2020).

Most of these problems will hopefully soon be in the past as long-read sequencing technologies and the experimental phasing of sex chromosomes eliminate many of the bioinformatic problems that plague numerous current studies.

ACKNOWLEDGEMENTS

We are grateful to members of the Kirkpatrick lab for discussion and comments. This study was supported by NIH grant R01-GM116853 to M.K..

REFERENCES

Almeida P, Sandkam B, Morris J, Darolti I, Breden F, Mank J (2020) Divergence and remarkable diversity of the Y chromosome in guppies. *BioRXiv* **doi.org/10.1101/2020.07.13.200196**.

Andrews S (2010) *FastQC: a quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice WR, Valenzuela N (2011) Are all sex chromosomes created equal? *Trends in Genetics* **27**, 350-357.

Bachtrog D, Mank J, Peichel C, Otto S, Kirkpatrick M, Ashman T-L, Hahn M, Kitano J, Mayrose I, Ming R, Perrin N, Ross L, Valenzuela N, Vamosi J (2014) Sex determination: Why so many ways of doing it? *PLoS Biology* **12**, e1001899.

Bergero R, Charlesworth D (2019) Reply to Wright *et al.*: How to explain the absence of extensive Y-specific regions in the guppy sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 12609-12610.

Bergero R, Gardner J, Bader B, Yong L, Charlesworth D (2019) Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 6924-6931.

Bissegger M, Laurentino TG, Roesti M, Berner D (2019) Widespread intersexual differentiation across the sickleback genome — the signature of sexually antagonistic selection? *Molecular Ecology* **29**, 262-271.

Brelsford A, Dufresnes C, Perrin N (2016) High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity* **116**, 177-181.

Broad Institute (2019) *Picard Toolkit*.

Charlesworth D (2017) Evolution of recombination rates between sex chromosomes. *Philosophical Transactions of the Royal Society B-Biological Sciences* **372**, 20160456.

Charlesworth D (2018) The guppy sex chromosome system and the sexually antagonistic polymorphism hypothesis for Y chromosome recombination suppression. *Genes* **9**, 264.

Charlesworth D, Bergero R, Graham C, Gardner J, Yong L (2020a) Locating the sex determining region of linkage group 12 of the guppy (*Poecilia reticulata*). *BioRXiv* **doi.org/10.1101/2020.04.30.063420**.

Charlesworth D, Zhang Y, Bergero R, Graham C, Gardner J, Yong L (2020b) Using GC content to compare recombination patterns on the sex chromosomes and autosomes of the guppy, *Poecilia reticulata*, and its close outgroup species. *BioRXiv* **doi.org/10.1101/2020.05.22.110635**.

Cheng C, Kirkpatrick M (2020) The signal of sex-specific selection in humans is not an artefact: Reply to Mank et al. *Molecular Ecology* **29**, 1406-1407.

Darolti I, Wright AE, Mank JE (2020) Guppy Y chromosome integrity maintained by incomplete recombination suppression. *Genome Biology and Evolution* **12**, 965-977.

Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farre M, Fuller RC, Bourne GR, Larkin DM, Breden F, Mank JE (2019) Extreme heterogeneity in sex chromosome differentiation and

dosage compensation in livebearers. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 19031-19036.

Dixon G, Kitano J, Kirkpatrick M (2018) The origin of a new sex chromosome by introgression between two stickleback fishes. *Molecular Biology and Evolution* **36**, 28-38.

Fraser BA, Kunstner A, Reznick DN, Dreyer C, Weigel D (2015) Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). *Molecular Ecology* **24**, 389-408.

Fraser BA, Whiting JR, Paris JR, Weadick CJ, Parsons PJ, Charlesworth D, Bergero R, Bemm F, Hoffmann M, Kottler VA, Liu C, Dreyer C, Weigel D (2020) Improved reference genome uncovers novel sex-linked regions in the guppy. *Genome Biology and Evolution* **(in press)**.

Guerrero RF, Kirkpatrick M, Perrin N (2012) Cryptic recombination in the ever-young sex chromosomes of hylid frogs. *Journal of Evolutionary Biology*, 1947-1954.

Haskins CP, Haskins EF, McLaughlin JJA, Hewitt RE (1961) Polymorphism and population structure in *Lebistes reticulatus*, a population study. In: *Vertebrate Speciation* (ed. Blair WF), pp. 320-395. University of Texas Press, Austin.

Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, Barrow LN, Canestrelli D, Crochet P-A, Dufresnes C, Fu J, Ma W-J, Garcia CM, Ghali K, Nicieza AG, O'Donnell RP, Rodrigues N, Romano A, Martínez-Solano Í, Stepanyan I, Zumbach S, Brelsford A, Perrin N (2018) A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications* **9**, 4088.

Kirkpatrick M, Guerrero RF (2014) Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* **197**, 531-541.

Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, Dreyer C (2016) The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. *Plos One* **11**, e0169087.

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Liao Y, Smyth GK, Shi W (2014) *featureCounts*: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930.

Lindholm A, Breden F (2002) Sex chromosomes and sexual selection in poeciliid fishes. *American Naturalist* **160**, S214-S224.

Lisachov AP, Zadesenets KS, Rubtsov NB, Borodin PM (2015) Sex chromosome synapsis and recombination in male guppies. *Zebrafish* **12**, 174-180.

Martin M (2011) Cutadapt removes adapter sequences. from high-throughput sequencing reads. *EMBnet Journal* **17**, 10-12.

Meisel RP (2020) Evolution of sex determination and sex chromosomes: a novel alternative paradigm. *Bioessays* **2020**, 1900212.

Morris J, Darolti I, Bloch NI, Wright AE, Mank JE (2018) Shared and species-specific patterns of nascent Y chromosome evolution in two guppy species. *Genes* **9**, 238.

Muyle A, Kafer J, Zemp N, Mousset S, Picard F, Marais GAB (2016) SEX-DETector: a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution* **8**, 2530-2543.

Nanda I, Schories S, Tripathi N, Dreyer C, Haaf T, Schmid M, Schartl M (2014) Sex chromosome polymorphism in guppies. *Chromosoma* **123**, 373-383.

Quinlan AR, Hall IM (2010) *BEDTools*: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

Rodrigues N, Studer T, Dufresnes C, Ma WJ, Veltsos P, Perrin N (2017) Dmrt1 polymorphism and sex-chromosome differentiation in *Rana temporaria*. *Molecular Ecology* **26**, 4897-4905.

Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M (2020) Contrasting tempos of sex chromosome degeneration in sticklebacks. *BioRXiv* **https://doi.org/10.1101/2020.09.17.300236**.

Sardell JM, Kirkpatrick M (2020) Sex differences in the recombination landscape. *American Naturalist* **195**, 361-379.

Schartl M, Walter RB, Shen YJ, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP, Bisazza A, Minx P, Hillier L, Wilson RK, Fuerstenberg S, Boore J, Searle S, Postlethwait JH, Warren WC (2013) The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature Genetics* **45**, 567-572.

Shen YJ, Chalopin D, Garcia T, Boswell M, Boswell W, Shiryev SA, Agarwala R, Volff JN, Postlethwait JH, Schartl M, Minx P, Warren WC, Walter RB (2016) *X. couchianus* and *X. hellerii* genome models provide genomic variation insight among *Xiphophorus* species. *BMC Genomics* **17**, 67.

Toups MA, Rodrigues N, Perrin N, Kirkpatrick M (2019) A reciprocal translocation radically reshapes sex-linked inheritance in the common frog. *Molecular Ecology* **28**, 1877-1889.

Tripathi N, Hoffmann M, Willing E-M, Lanz C, Weigel D, 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 of London. Series B: Biological Sciences* **276**, 2195-2208.

van Doorn GS, Kirkpatrick M (2007) Turnover of sexual chromosomes induced by sexual conflict. *Nature* **449**, 909-912.

Vicoso B (2019) Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology & Evolution* **3**, 1632-1641.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Winge O (1922) One-sided masculine and sex-linked inheritance in *Lebistes reticulatus*. *Journal of Genetics* **12**, 145-162.

Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE (2017) Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nature Communications* **8**, 14251.

Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam BA, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE (2019) On the power to detect rare recombination events. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 12607-12608.

DATA ACCESSIBILITY

All the data used in the study were acquired from the publicly-accessible sources cited in the text. The custom scripts used for data processing, analysis, and figure generation are available on Github at https://github.com/grovesdixon/guppy_sex_chroms.

AUTHOR CONTRIBUTIONS

M.K. conceived of the study. G.D. and J.M.S. conducted the analyses. All authors contributed to the writing.

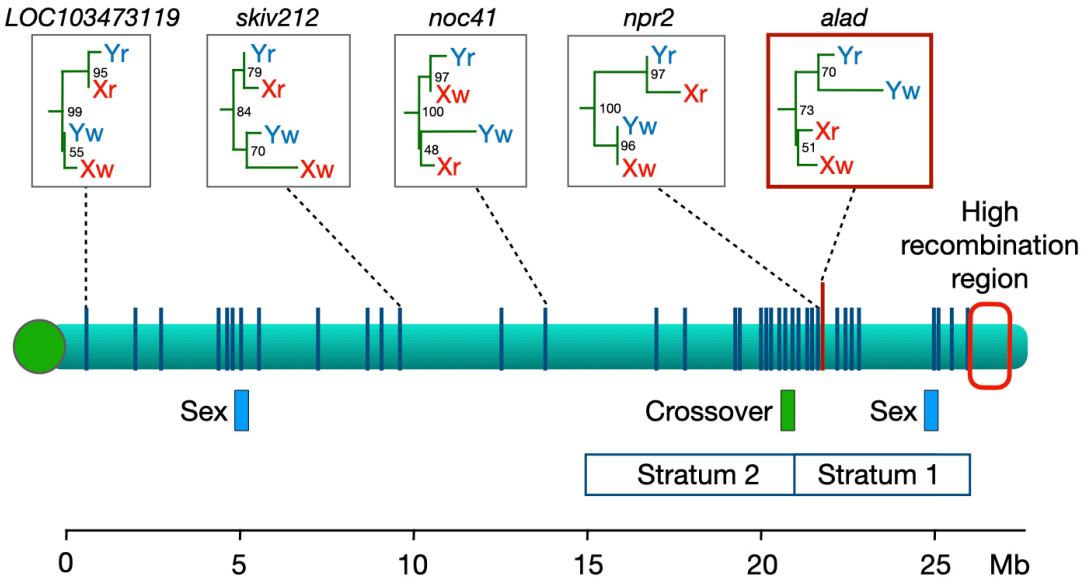

Fig 1

Figure 1: The sex chromosome of the guppy, *P. reticulata*. The horizontal green bar shows the interval in which a crossover was observed by Bergero *et al.* (2019; D. Charlesworth pers. comm.). Stratum 1 and Stratum 2 are regions where Wright *et al.* (2017) reported that the X and Y do not recombine. Blue boxes labeled "Sex" are candidate regions for the male-determining factor (Fraser *et al.* 2020). At far right is a region with a high local recombination rate (Lisachov *et al.* 2015, Bergero *et al.* 2019, Darolti *et al.* 2019). Vertical blue lines show locations of the 42 loci at which gene trees for sequences from *P. reticulata* (Xr, Yr) and *P. wingei* (Xw, Yw) were estimated by Darolti *et al.* (2020). Only the gene tree highlighted in the red box has a topology consistent with an ancestral SDR shared by the two species; examples of four other representative gene trees are also shown. Numbers at their nodes give the bootstrap support. The centromere is shown as the green circle at left.

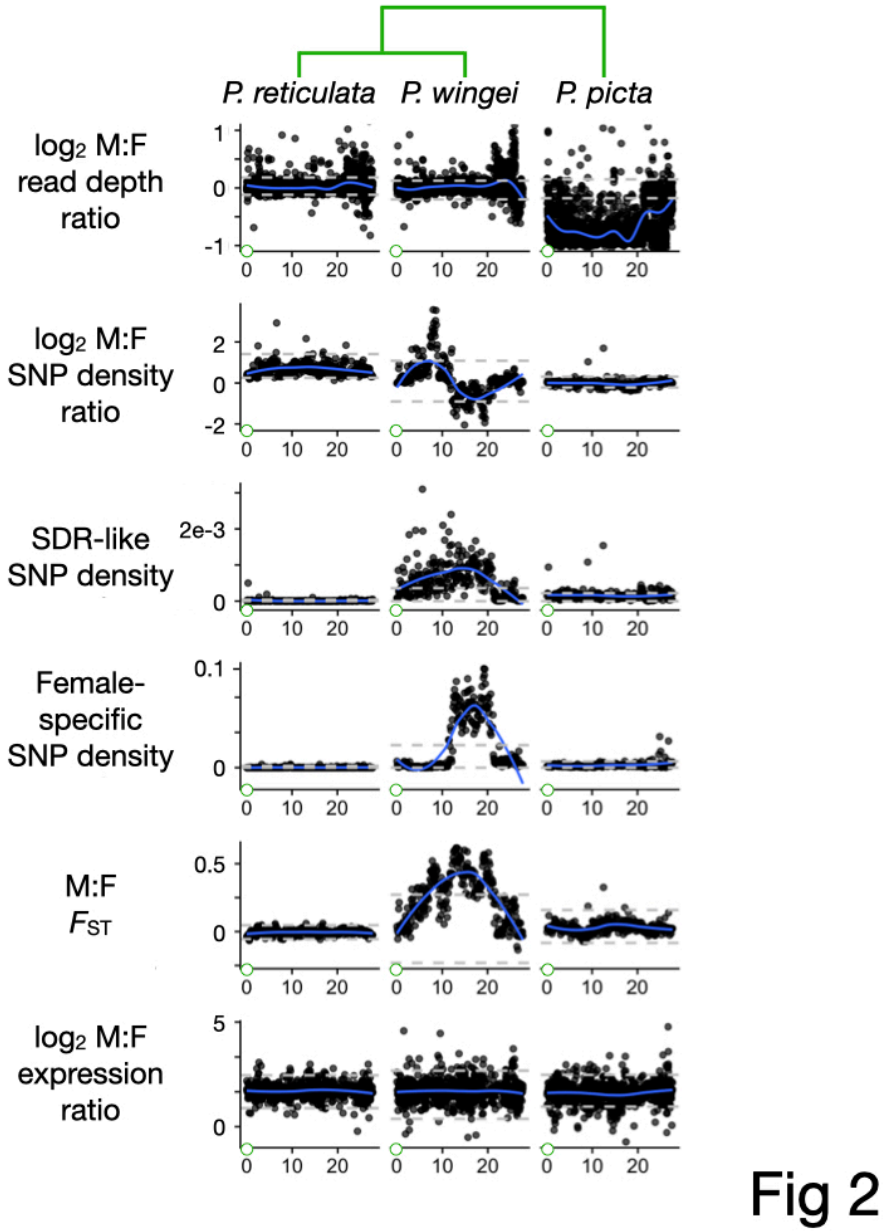

Figure 2. Divergence between the X and Y chromosomes in the three species as reflected by six statistics that measure differences between males and females. Points are averages for sliding windows of 10 kb (read depth ratio) or 100 kb (all other statistics). The gray horizontal dashed lines show the bottom 2.5% and top 97.5% intervals based on windows from all autosomes. The blue curves are smoothed regressions. Green circles at the left of the Y axes represent the centromere. A schematic of the phylogeny is shown at top.

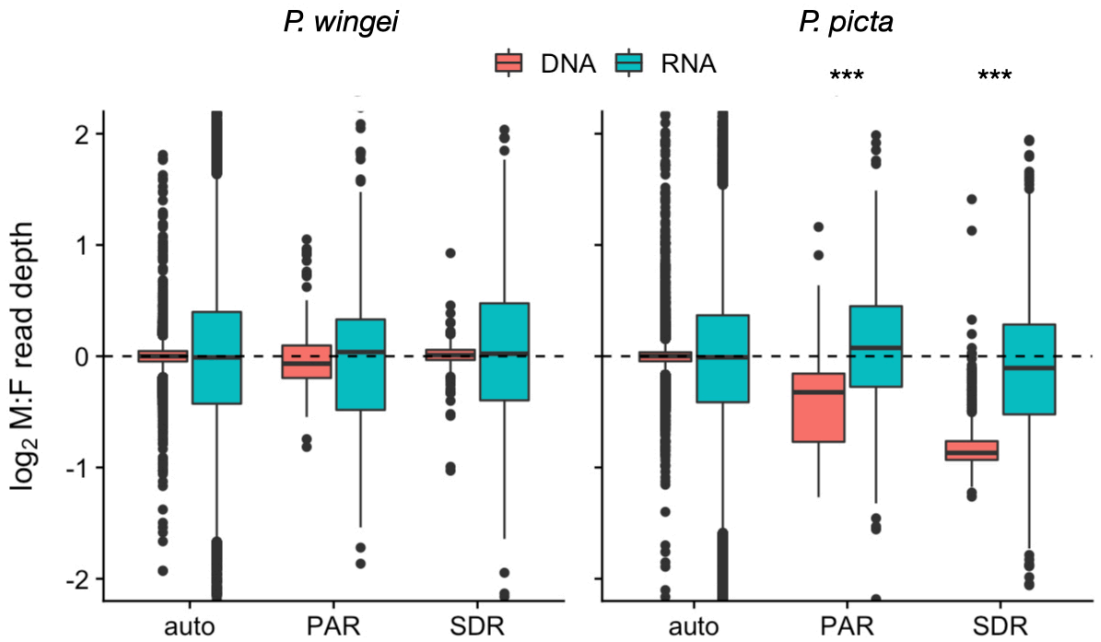

Fig 3

Figure 3. The ratio of male:female read depth (expressed as $\log_2$) at coding loci on the autosomes (auto), pseudoautosomal region (PAR), and sex determining region (SDR) for both genomic DNA and RNA transcripts. In *P. picta*, read depth for genomic DNA on the sex chromosomes in males is about half of what it is in females, consistent with most of the Y being deleted. The RNA transcripts are about equally abundant in the two sexes, however, indicative of dosage compensation. Results for *P. reticulata*, which are omitted for visual clarity, are very similar to *P. wingei*. *** The difference between the DNA and RNA read depth ratios in the PAR and SDR are significantly greater in *P. picta* than in *P. wingei* ($p << 10^{-13}$, one-tailed paired *t*-test).

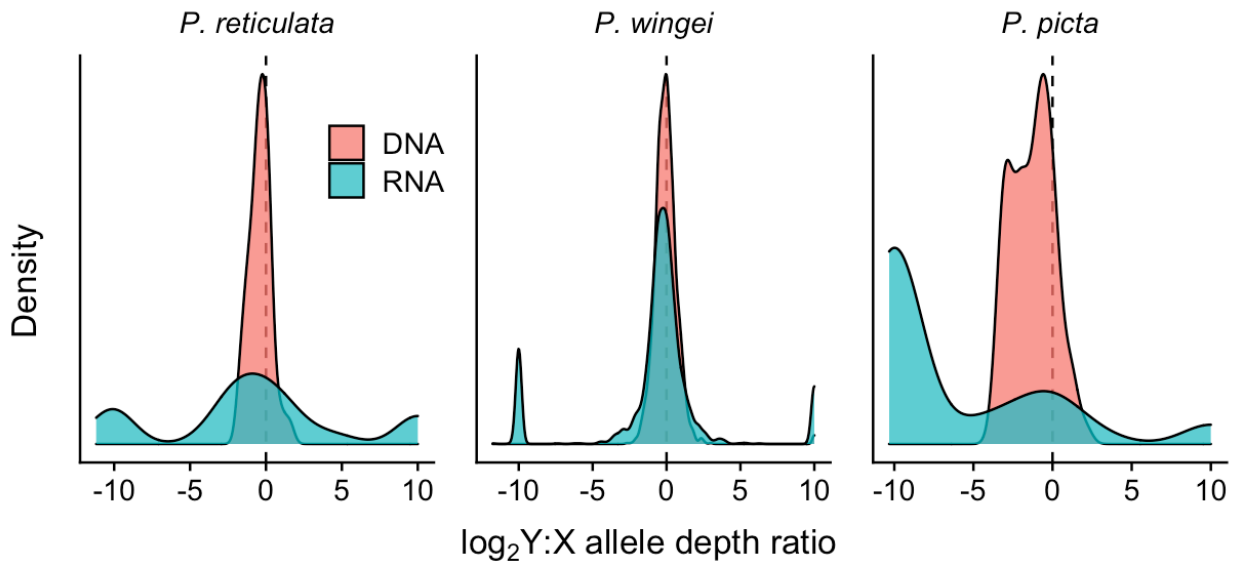

Figure 4. The distributions of the ratios of read depth (expressed as $\log_2$) for Y-like and X-like alleles in males for both genomic DNA and RNA transcripts. When no RNA reads were recovered from either the Y-like or X-like allele, the ratio was set to -10 or 10, respectively. Many Y-like alleles of *P. picta* have little or no expression, consistent with the interpretation that dosage compensation has occurred, resulting in the equal expression of most of the genes on the sex chromosomes (Figure 3).

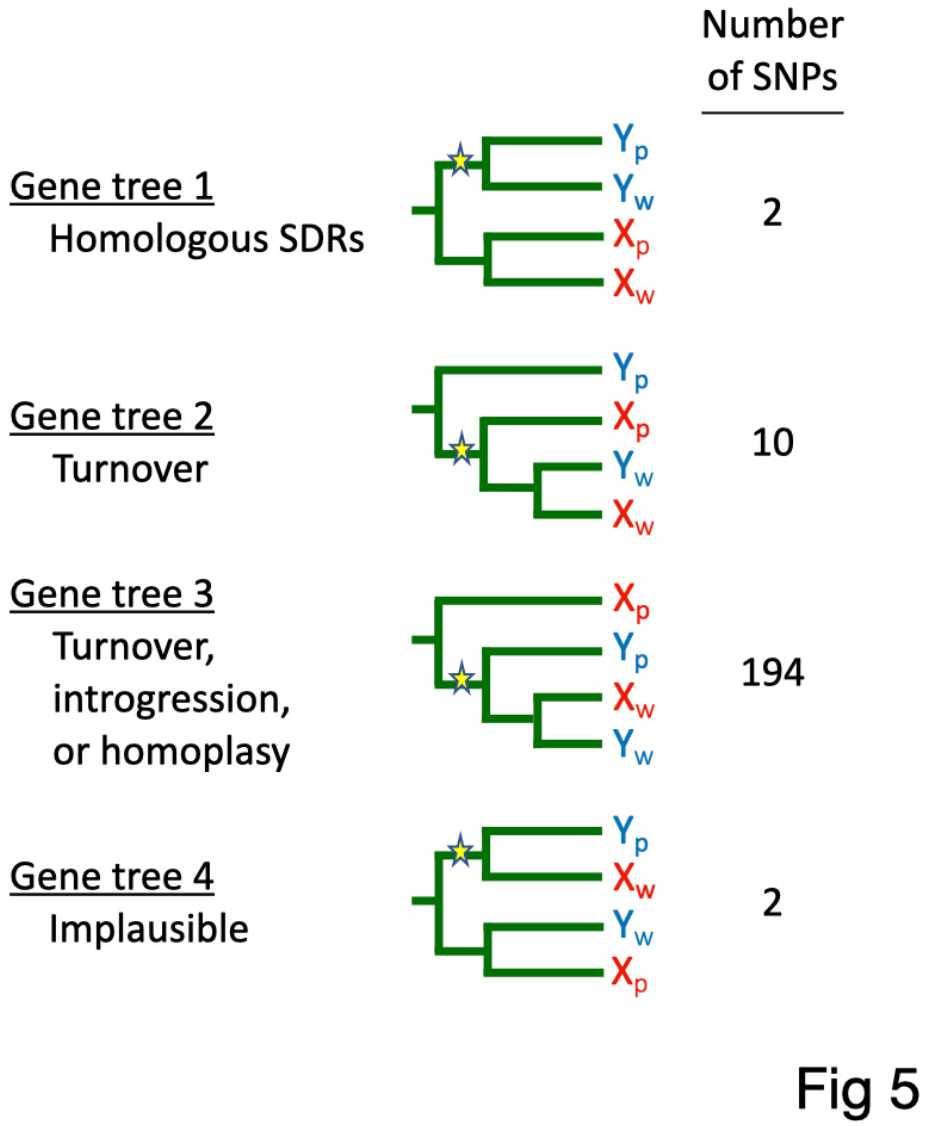

Figure 5. Topologies of four gene trees estimated from SNPs in the SDRs of *P. picta* and *P. wingei*. X_p and Y_p represent the sex chromosomes of *P. picta*; X_w and Y_w represent the sex chromosomes of *P. wingei*. Derived mutations are shown by stars. At right are the number of SDR-like SNPs consistent with each topology. Tree 1 is consistent with homologous SDRs in the two species. Tree 2 is consistent with a turnover in which the *P. wingei* Y was derived from an X. Tree 3 is consistent with a turnover in which the *P. wingei* X was derived from a Y, introgression of a Y from the *wingei-reticulata* lineage into *P. picta*, or introgression of the *P. picta* X from an outgroup species. Tree 4 is biologically implausible.

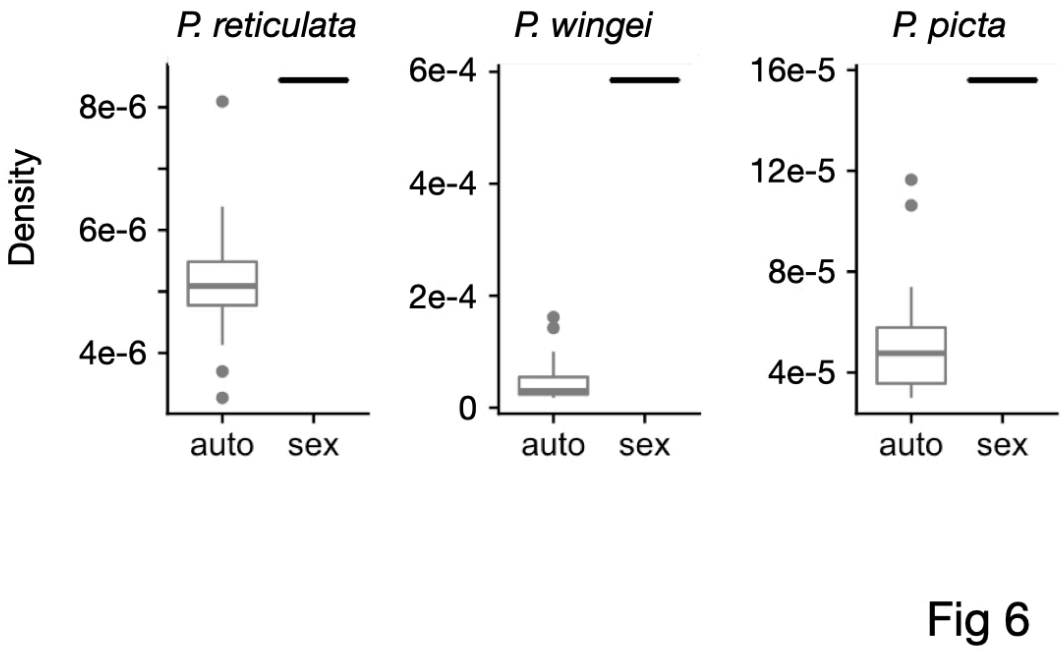

Figure 6. Density per base pair of SDR-like SNPs for the autosomes and sex chromosomes.