Partial sex linkage and linkage disequilibrium on the guppy sex chromosome

Received: 2 March 2022 | Revised: 1 August 2022 | Accepted: 9 August 2022

DOI: 10.1111/mec.16674

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Partial sex linkage and linkage disequilibrium on the guppy sex chromosome

Suo Qiu[1] | Lengxob Yong[2,3] | Alastair Wilson[2] | Darren P. Croft[4] | Chay Graham[1] | Deborah Charlesworth[1]

[1]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

[2]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, UK

[3]Marine Resources Research Institute, South Carolina Department of Natural Resources, Charleston, South Carolina, USA

[4]Centre for Research in Animal Behaviour, College of Life and Environmental Sciences, University of Exeter, Exeter, UK

Correspondence
Deborah Charlesworth, Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3LF, UK.
Email: deborah.charlesworth@ed.ac.uk

Funding information
H2020 European Research Council, Grant/Award Number: 695225

Handling Editor: Nick Barton

Abstract

The guppy Y chromosome has been considered a model system for the evolution of suppressed recombination between sex chromosomes, and it has been proposed that complete sex-linkage has evolved across about 3 Mb surrounding this fish's sex-determining locus, followed by recombination suppression across a further 7 Mb of the 23 Mb XY pair, forming younger "evolutionary strata". Sequences of the guppy genome show that Y is very similar to the X chromosome. Knowing which parts of the Y are completely nonrecombining, and whether there is indeed a large completely nonrecombining region, are important for understanding its evolution. Here, we describe analyses of PoolSeq data in samples from within multiple natural populations from Trinidad, yielding new results that support previous evidence for occasional recombination between the guppy Y and X. We detected recent demographic changes, notably that downstream populations have higher synonymous site diversity than upstream ones and other expected signals of bottlenecks. We detected evidence of associations between sequence variants and the sex-determining locus, rather than divergence under a complete lack of recombination. Although recombination is infrequent, it is frequent enough that associations with SNPs can suggest the region in which the sex-determining locus must be located. Diversity is elevated across a physically large region of the sex chromosome, conforming to predictions for a genome region with infrequent recombination that carries one or more sexually antagonistic polymorphisms. However, no consistently male-specific variants were found, supporting the suggestion that any completely sex-linked region may be very small.

KEYWORDS
balancing selection, evolutionary strata, genome assembly, linkage disequilibrium, partial sex linkage, sexual antagonism

1 | INTRODUCTION

Genetic sex determination and sexual dimorphism were first studied in the guppy, *Poecilia reticulata* (formerly *Lebistes reticulatus*) just over 100 years ago (Schmidt, 1920), and the species has been considered a good system for understanding the lack of recombination between sex chromosome pairs. This fish has male heterogamety, though like the sex chromosomes of many other fish, the XY pair is homomorphic, or, at most, slightly heteromorphic, with the Y being larger than the X (Lisachov et al., 2015; Nanda et al., 2014; Traut & Winking, 2001). Recent genome sequence analyses have shown that the gene contents of this chromosome pair are very similar (Bergero et al., 2019; Charlesworth et al., 2020; Darolti et al., 2020; Fraser et al., 2020; Wright et al., 2017), unlike the situation in species such as mammals and birds (with sex chromosome heteromorphism and loss of many genes from the completely nonrecombining Y chromosomal regions). It has nevertheless been proposed that recombination suppression has evolved in guppies, producing extensive nonrecombining regions, or "evolutionary strata" (Wright et al., 2017). Strata are nonrecombining regions with distinctively different Y–X sequence divergence levels detected in heteromorphic sex chromosome pairs, including those of humans and other mammals (Cortez et al., 2014; Lahn & Page, 1999), birds (Xu et al., 2019) and the plant *Silene latifolia* (Papadopulos et al., 2015).

It is important to understand whether the guppy XY pair has evolutionary strata because their presence implies that recombination suppression evolved, rather than alternative possibilities. One possibility is that a sex-determining locus evolved within an already nonrecombining genome region, and another is that a male-determining factor duplicated into a sex chromosome, in a turnover event, hindering pairing and directly stopping recombination (Charlesworth, 2019; Charlesworth et al., 2021). The most plausible situation generating selection for recombination suppression is the presence of sexually antagonistic (SA) polymorphisms closely linked to the sex-determining locus (Rice, 1987, recently reviewed by Dagilis et al., 2022). Trinidadian guppy populations are suitable for testing this idea, because their Y chromosomes carry SA male coloration factors that are advantageous in males, and increase male mating success, outweighing the disadvantage of being more conspicuous to predatory fish that are present along with guppies in their natural habitats (Endler, 1980; Houde, 1992). These factors are polymorphic within Trinidadian guppy populations (Haskins et al., 1961).

Under any of the possibilities just outlined, the presence of the sex-determining locus defines an oldest stratum (though this could be confined to a single sex-determining SNP). The guppy sex-determining locus is thought to be located roughly 30% of the physical distance from the telomere of the roughly 26.5 Mb telocentric XY pair, chromosome 12, or LG12 (Lisachov et al., 2015; Nanda et al., 2014; Tripathi et al., 2009). An old stratum was detected between 22 and 25 Mb in the assembly of LG12, based on a sequence coverage analysis that suggested slightly diverged Y-linked sequences. Younger strata occupying about 7 Mb of the more centromere-proximal part of the XY pair were suggested by slightly elevated SNP density in males from natural populations, compared with that in two sequenced females (Darolti et al., 2020; Wright et al., 2017). However, these regions are unlikely to be completely nonrecombining, as the guppy Y crosses over occasionally with the X chromosome, so that most of the chromosome is partially, rather than completely, sex-linked (reviewed by Lindholm & Breden, 2002). Nevertheless, recombination rates are low across most of the chromosome, and molecular markers only rarely detect crossovers proximal to about 25.5 Mb (Bergero et al., 2019; Charlesworth et al., 2020).

Although studies of guppy samples consistently yield evidence that Y-linked sequences differ from their X-linked counterparts, divergence between Y and X sequences is minimal, and clearly defined strata containing many genes, are not detected. A balanced polymorphism generates associations of closely linked neutral variants, and affects diversity similarly to population subdivision, leading to a local peak of differentiation in the region. This is potentially detectable using F_{ST} analysis, which estimates the proportion of total diversity that reflects differences between sequences in two populations, and is equivalent to linkage disequilibrium (LD) between the balanced polymorphism and the neutral variants (Charlesworth et al., 1997). F_{ST} between the sexes can therefore be used to quantify differences between sequences in male and female populations reflecting LD with the male-determining locus. Previous studies describe hints that associations between sequence variants and the sex-determining locus are commonest in the region distal to 20 Mb of the guppy sex chromosome, but no completely male-specific variants (indicating complete Y-linkage) have yet been found (Bergero et al., 2019; Charlesworth et al., 2020; Darolti et al., 2020; Fraser et al., 2020; Wright et al., 2017). Here, we describe new evidence supporting the possibility that linkage disequilibrium (LD) in this rarely recombining chromosome, rather than recent recombination suppression creating new strata, has shaped diversity across the guppy XY pair.

In a population of constant effective population size N_e, the diversity at neutral sites linked with a selected locus depends on the quantity $\rho = 4N_e r$, where r is the recombination rate between the two loci. Figure 1 illustrates that the effects on diversity in both sexes near a sex-determining locus is restricted to a physically very small region, unless N_e is very small. Kirkpatrick and Guerrero (2014) showed that a polymorphic SA factor present in a partially sex-linked region closely linked to a sex-determining locus enlarges the physical distances across which neutral variants show associations with individuals' sexes. This effect might be detectable in the guppy, as, if the Y rarely recombines with the X in male meiosis, diversity in males could be high across a physically large region. This theory also shows that, if recombination occurs often enough, clusters of variants associated with the sexes could potentially identify the locations of the sex-determining and/or SA loci. A recent study inferred multiple SA polymorphisms in a recombining region of a stickleback neo-sex chromosome (Dagilis et al., 2022). Although small sample sizes create chance associations, even between sequence variants that are not closely linked (Park, 2019), large enough samples to distinguish

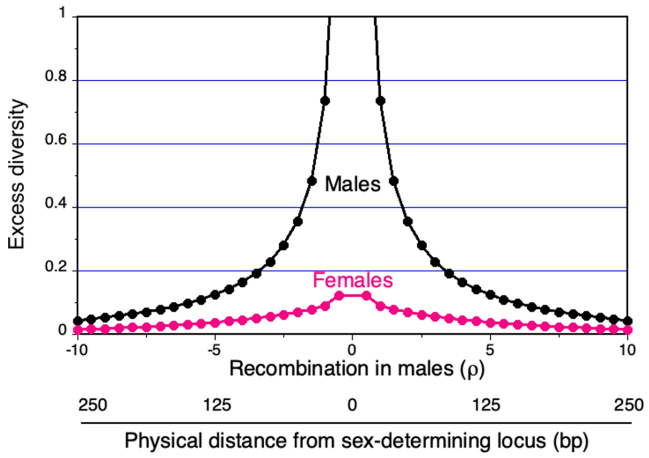

FIGURE 1 Peak of diversity predicted in males and females at regions linked to a male-determining factor for the case with no sexually antagonistic polymorphisms in the region. The values are based on deterministic equations for coalescence times between pairs of X-linked sequences and between Y- and X-linked ones (Kirkpatrick et al., 2010). The elevated Y-X divergence times can be detected in phased data as higher Y-X divergence values at neutral sites close to the male-determining factor, or in unphased data as higher diversity in males than females. Coalescence times and diversity values are also elevated for X chromosome sequences, but only slightly. The *y*-axis shows the excess in coalescence time (and thus of neutral diversity) at different distances from a male-determining factor, compared with unlinked sites. Distances are shown on the *x*-axis as the recombination rate in males, ρ, or $4Nr$ values, where r is the recombination distance in Morgans, and also as the physical distances assuming a very low recombination rate of 0.01 cM/Mb and an N value of 1 million. Higher recombination rates decrease the physical distances proportionately.

such associations from complete sex linkage (evolutionary strata), can now be obtained.

To understand associations between sex and chromosome 12 variants in populations of Trinidadian guppies, the demographic history of the populations sampled is important, particularly for upstream populations, which probably underwent bottlenecks when these sites were colonized (Magurran, 2005). Bottlenecks cause loss of low frequency variants, reducing diversity but increasing the variance in diversity values. Neutral variants in regions of the sex chromosome pair distant from the sex-determining locus could thus appear as outliers with high F_{ST} values between the sexes. This may explain high chromosome-wide F_{ST} values occasionally found between males and females, including a Guanapo river low predation population sample (Fraser et al., 2020), and a captive sample from an Aripo river high predation site (Bergero et al., 2019). On the other hand, male-specific variants (creating higher sequence diversity in males than females near an XY system's male-determining locus, as illustrated in Figure 1) might be clearest in bottlenecked populations, with low diversity elsewhere in the genome.

Populations of guppies with the SA polymorphisms mentioned above are present in most rivers in the Northern Range mountains of Trinidad. Some of the polymorphisms have probably been maintained for long times, as they are shared by many different populations (Haskins et al., 1961). They are found in both upstream sites, under low predation pressure, and downstream ones, with heavy predation (reviewed in Magurran, 2005, and in Tobago Haskins et al., 1961). Their maintenance involves rare male advantages (Fraser et al., 2013; Hughes et al., 2013; Olendorf et al., 2006), not solely SA selection. Nevertheless, these polymorphisms could have created selection favouring suppressed recombination, perhaps transiently, since the male coloration factors found in nature are male-limited in expression (reviewed in Haskins et al., 1961). Wright et al. (2017) indeed concluded that recombination was suppressed across large parts of the guppy LG12, after an older nonrecombining region carrying the sex-determining locus evolved, and that such changes occurred only in upstream, low predation sites.

The effect of different predation regimes on the maintenance of partially sex-linked SA coloration factors on the evolution of suppressed recombination have not been modelled, and it cannot be ruled out that the ecological differences between up- and downstream sites reviewed by Magurran (2005) might allow sex chromosome strata to evolve only in upstream sites. For example, in downstream sites, selective disadvantages to recombinant females that inherit and express coloration factors might keep the frequency of these factors so low that selection against recombinants might be too weak to select for closer linkage with the male-determining factor. There is, however, direct evidence for more frequent (though still rare) recombination between a sex-linked male coloration factor (*Sb*) and the male-determining locus in up- than downstream males from the Aripo river (Haskins et al., 1961). Moreover, stronger associations between the male-determining locus and coloration factors in high- than low-predation males were confirmed in other rivers, using testosterone treatment to reveal females carrying *Sb* (Haskins et al., 1961), and by consistent results in other rivers for other coloration factors (Gordon et al., 2012). Overall, these results support the view that low recombination rates between coloration factors and the male-determining locus prevail in high-predation conditions, but that, after colonization of upstream sites, with weaker predation pressure, recombinants are less disfavoured (reviewed in Charlesworth, 2018). If so, close linkage should be reversible after guppies colonize low-predation sites, which is consistent with changes detected in evolutionary experiments (Gordon et al., 2017).

The conclusion that new evolutionary strata have evolved in low-predation populations is further called into question by evidence for occasional crossovers in the regions identified as younger strata (Almeida et al., 2021; Bergero et al., 2019; Charlesworth et al., 2020; Darolti et al., 2020). Even rare crossing over prevents accumulation of completely Y-specific variants (see Figure 1), and indeed these studies found no consistently male-specific variants in guppy populations. It is therefore worth considering whether randomly generated associations in bottlenecked populations can account for associations between sequence variant and the sexes across large regions of chromosome 12, especially in low-predation populations.

Our goals in the present study were: (1) To analyse genome-wide patterns of synonymous site diversity to test for recent bottlenecks in upstream Trinidadian guppy populations. (2) To compare diversity

between the sexes for the XY pair and the autosomes in our natural population samples. Autosomal diversity was previously estimated by Almeida et al. (2021), but only F_{ST} values between the sexes and relative X/A and Y/A values were reported. However, it is important to consider absolute diversity values, because high F_{ST} values can be due to low diversity, as they estimate the proportion of total diversity that is found between, rather than within populations, and low diversity within one or both of the populations being compared necessarily implies high F_{ST} (Charlesworth, 1998; Charlesworth et al., 1997; Cruickshank & Hahn, 2014). In the case of F_{ST} values between the sexes, an advantageous X-linked mutation that has recently spread through a population, causing a selective sweep, will result in low X chromosome diversity, creating locally high male–female F_{ST}. In contrast, the hypothesis of LD due to close linkage to a locus under balancing selection (as in Figure 1) predicts that the region where the guppy male-determining factor is located (or a wider region that recombines extremely rarely in males) should have high nucleotide diversity in males, compared with other genome regions in individuals sampled from the same population. In the same region, diversity should also be slightly elevated in females, as shown in Figure 1. (3) To examine the sex chromosome region in which SNP genotypes show associations with individuals' phenotypic sexes, to narrow down the location of the guppy male-determining factor, and to assess the likely size of any nonrecombining region, or evolutionary stratum. In the Discussion section, we discuss problems that may be responsible for the difficulty experienced in locating this factor.

2 | MATERIALS AND METHODS

2.1 | Trinidadian guppy samples

Fish were sampled from 12 natural populations in Trinidad (Table S1), and 20 males and 20 females from each population were preserved in the field (with approval by Trinidad's Ministry of Agriculture, Land and Fisheries) before transport to the UK for DNA extraction and sequencing.

2.2 | Mapping pooled sequences to guppy female assembly

Low coverage Illumina sequencing was done using NovaSeq 6000 machine, whose error rate is less than 1% (Stoler & Nekrutenko, 2021). Separate male and female pools, each of 20 individuals, were sequenced for each sample, yielding sequence lengths of 150 bp, and the raw sequence reads from the 24 pools were processed as described in detail previously (Yong et al., 2021). Briefly, reads were trimmed to remove low quality bases and adapter sequences using CUTADAPT. The cleaned reads were mapped to the guppy female assembly version 1.0 (accession number: GCF_000633615.1) using BWA (Li & Durbin, 2010). Mapped reads were sorted and PCR duplicates were removed using SAMTOOLS version 1.9 (Li et al., 2009). Reads around indels were realigned using the modules RealignerTargetCreator and IndelRealigner implemented in gatk version 3.4. Finally, the input file for the diversity analyses described in the next section was generated using the Samtools mpileup function, reads with mapping quality lower than 20 were discarded.

2.3 | Nucleotide diversity analyses in males and females

To ensure that the sequences used are reliably aligned single-copy genes, and the variants analysed are likely to behave close to neutrally, we used synonymous sites in annotated genes for population genomic analyses of nucleotide diversity. The guppy genome GFF3 file from Ensembl (release 97, https://www.ensembl.org/pub/release-95/gff3/poecilia_reticulata/) provided the annotation details required. Synonymous nucleotide diversity (π_S) and Watterson's Θ_S were estimated using the POPOOLATION software (Kofler et al., 2011); the software's Syn-nonsyn-sliding.pl and Syn-nonsyn-at-position.pl scripts were used to estimate these quantities in nonoverlapping 250 kbp windows and in individual genes, which yielded similar results (see below). To obtain reliable diversity estimates, SNPs with coverage lower than one third of the average coverage for the sample, or higher than three times the average coverage, were first filtered out. SNPs for these calculations were defined with a minimum number of reads supporting the minor allele, *mc*, of either 1, which yielded π_S results very similar to those using the SNPGENIE software (Nelson et al., 2015), or 2, as suggested for data with low error rates like those for our sequencing, based on simulations (Kofler et al., 2011). Because Θ_S estimates depend on rare variants (the majority of variants in natural populations), the threshold in PoPoolation analyses affects these values, but the Results section shows that the conclusions were unaffected.

Diversity values were estimated separately for the samples of males and females from each collection site. Finally, values of the quantity $1-(\pi_S/\Theta_S)$, denoted by $\Delta\Theta_S$, were calculated for each window or gene. This detects frequency differences between variants, as values of Θ are estimated using variants irrespective of their frequencies, whereas π values largely reflect intermediate frequency variants. $\Delta\Theta_S$ can therefore detect the excess of rare variants that is expected in a population as it returns towards the neutral equilibrium frequency distribution after a recent bottleneck; in such populations, higher values of $\Delta\Theta_S$ are expected (Tajima, 1989). The $\Delta\Theta_S$ measure is preferable to Tajima's *D* values as it is less affected by the lengths of the sequences analysed.

2.4 | PoolSeq detection of candidate fully sex-linked sites and sex-linked regions

To test for candidate completely sex linked variants on the sex chromosome pair, we screened the data from each sample of 20 males and 20 females to detect sites with variants consistently showing

the genotype configuration expected under complete sex linkage: under male heterogamety, the guppy sex-chromosome system (Winge, 1922), all females will be homozygous for an X-linked variant at candidate sites, while all males will be heterozygous for a different variant. Male-specificity in a large enough sample indicates Y linkage, though not necessarily complete Y linkage (see discussion below and Figure 1). Sites with the configuration suggesting XX in females and XY in males are called "XY sites" in what follows. This analysis used all site types, not just synonymous sites, because variants of any kind can help detect sex linkage. Previous analyses using Illumina sequencing data in the guppy (Almeida et al., 2021; Bergero et al., 2019; Charlesworth et al., 2020) and a *Nothobranchius* species (Reichwald et al., 2009) also analysed regional differences in the proportions of polymorphic sites with genotype configurations suggesting different strengths of association with sex-determining loci, but sample sizes were small, or the individuals were from captive populations. Large samples from multiple natural populations are needed to narrow down the location of the guppy sex-determining locus, and our study adds new samples to those analysed previously, and describes the results of the analyses in greater detail, revealing valuable new information (see Section 3).

Table 1 illustrates our approach for such screening of PoolSeq data, based on the frequencies of alternative bases at biallelic sites. At a site where a variant was detected in a sample of females, the proportion of the variant (XFREQ) was computed using the total number of reads of either base, after the filtering described above (TOTAL). This detects candidate sites with no variation in the X chromosomes present in the females. At all such sites with variants detected in the male sample, we calculated the proportion of the variant, multiplied by 2, yielding YFREQ, the variant's estimated frequency in the population of Y chromosomes. The table illustrates the method with three sites. When all 20 males are heterozygotes for a base that differs from the one found in all females, the YFREQ value is 1, suggesting fixation of the variant in the Y population (as expected for the male-determining locus, which must be present in all males). YFREQ values of 1, suggesting fixation in the Y population, may also be found at sites completely linked to the male-determining factor.

Because the individuals were sequenced in pools of males and females, the true numbers will be less than 20 of each sex. It is nevertheless unlikely that many variants will be at high intermediate frequencies in the males, and have this YFREQ value, but yet be invariant in the female sample. For example, consider a site with both XFREQ and YFREQ values of 1, which thus has an estimated frequency of the rarer variant (detected only in the males) of 0.25 (see Table 1). Most sites satisfying this criterion will be XY sites that are likely to be sex linked, even if many fewer than 20 females were in fact sequenced. For instance, if half of the individuals of either sex carry the allele detected in the males, the binomial probability that the variant is not present in a female pool of only 10 individuals is 0.098% (ignoring genotyping errors in the male sample, or failure to detect the variant in heterozygous females). A further problem is that sequencing or mapping errors (or rare recombination events) might produce XFREQ values slightly below 1 in some samples, which would not be included as candidates for having male-specific variants. We therefore repeated the analysis with XFREQ ≥0.975 (i.e., one of the 40 alleles in our female sample matching the putatively Y-specific allele). As described below, this did not increase the sharing of candidate fully sex-linked sites between populations.

This approach has the potential to detect candidate sites with patterns suggesting complete sex linkage, and chromosome regions that show enrichment for such sites. Importantly, some variants at sites within a completely sex-linked region may have YFREQ values <1, because a recent mutation in a Y-linked sequence may not have become fixed at the site (although, in the absence of recombination, such a segregating variant will still be male-specific, and the site will have XFREQ = 1). Segregating male-specific variants might also be due to maintenance of different Y haplotypes. We consider this in the Discussion section, along with comparing this approach with alternative methods used for screening for sex-linkage. Variants in partially sex-linked regions may also have high YFREQ values, but both these, and their XFREQ values should be <1.

3 | RESULTS

3.1 | Nucleotide diversity in populations and evidence for recent population size bottlenecks

With both diversity measures (using the threshold *mc* value of 2 reads supporting the minor allele, which yields lower diversity values than *mc* = 1, see Section 2), most upstream (low-predation,

TABLE 1 Examples of the method for finding candidate fully sex-linked variants among variants detected in the sample from a natural population of a species known to have a genetic sex-determining system with male heterogamety.

Site		Counts of A:T:C:G at a site			
Position in the female assembly	REF base	Females	Males	XFREQ = X/TOTAL	YFREQ = Y/TOTAL × 2
12	T	0:**33**:0:0	0:**29**:0:1	1	0.017
9320	T	2:**44**:0:0	23:**48**:0:0	0.957	0.648
9414	G	14:0:0:**88**	105:0:0:**75**	0.863	1.167

Note: The commonest variant in the females (in bold font in the table), is defined as the X-linked allele and the alternative variant at the same site is inferred to be Y-linked. The total coverage at a site is denoted by "TOTAL".

or LP) populations have lower values than the downstream (high-predation, HP) ones from the same river, as expected (Figures 2 and 3), for both LG12 and autosomal sequences (except for the Quare river samples, individually labelled in Figure 3, with no diversity difference); the patterns are similar for diversity based on genes, rather than windows (Table S3A,B). Even including the Quare river, median π in upstream populations is at least 45% lower than in downstream ones, based on autosomal sequences from both sexes; this diversity measure is conservative for detecting diversity loss, as π depends most strongly on variants at intermediate frequencies; the differences in θ are larger, as this diversity measure is more weighted to low frequency variants (see Section 2), which are most likely to be lost in bottleneck events.

Both high- and low-predation samples also show deficits of low-frequency variants ($\Delta\theta_s$ values, see Section 2, are shown in Figure 3 and Table S3A,B). Because the sequences were obtained by PoolSeq, absolute values of this quantity cannot be interpreted, but the values are consistently higher in three of the four rivers where up- and downstream samples could be compared, with significant differences in both sexes, for both autosomal and sex-linked variants (Table S4). The results from three rivers with only one predation regime represented (Guanapo, Paria and Petit Marianne, see Table S1) are also consistent with this pattern (Figure 3, Table S3). However, the Quare upstream sample (Quare LP in Figure 3) is again unusual, as its unexpectedly high diversity (for both LG12 and autosomal sequences) is accompanied by an absence of evidence for loss of rare variants, and its $\Delta\theta_s$ value is low, like the values for downriver samples, indicating abundant rare variants. Both *mc* values yielded much higher $\Delta\theta_s$ values for upstream than downstream samples (though, as expected, θ_s is greatly reduced with *mc* = 2, rather than 1, see Table S3A,B). This consistent pattern cannot be due to erroneous inclusion of variants due to our chosen *mc* value.

Lower diversity of upstream samples is expected to lead to comparisons involving such samples having the highest F_{ST} values. This can explain the high values in comparisons between pairs of upstream populations from different rivers, using all site types, and lower ones between upstream and downstream populations from the same river (Fraser et al., 2015; Suk & Neff, 2009; Willing et al., 2010); the same is seen for the samples and synonymous sites studied here, using the PoolFstat package (Hivert et al., 2018; see Table S2).

3.2 | Male–female differentiation

We also detect higher synonymous site diversity in LG12 sequences in males than females, especially in the high diversity downstream samples (Figure 2 shows one river, as an example, Figure S2 shows results for all 12 populations sampled, and Figure S3 summarizes the mean values). Because separate pools were sequenced from males and females, and sex differences in diversity were not confined to the sex chromosome (Table S3A–C), we calculated sex chromosome/autosomal diversity ratios in each sex. Almost all samples have excess diversity on the sex chromosome (Figure 4 shows results with *mc* = 2, and Table S3A includes a plot with the results using *mc* = 1). Interestingly, this excess is detected in both sexes, not just in the males, though it is larger in males than females in nine of the 12 samples. In all high-predation populations, and in the low-predation Quare sample, excess diversity was detected in males across much of LG12; all samples had at least 70 windows with ratios above 1 (out of the total of 106,250 kb windows), highly significantly exceeding the expected 53 windows under the null hypothesis that this chromosome has diversity values similar to those for autosomal sequences. Among the female samples, however, only the high-predation one from the Guanapo river yielded a significant excess of such windows (69/106).

As the LG12 sequences sampled from females have at least as high diversity as autosomal sequences, our data do not suggest that the X chromosome has experienced recent hard selective sweeps creating low female-specific diversity regions. Genetic maps of LG12 in female guppies are consistently around 50 centiMorgans (Bergero et al., 2019; Charlesworth et al., 2020), implying that more than a single crossover event per meiosis is rare. A recent strong selective sweep event would therefore create a wide region of very low diversity, which is not seen (Figure S2B).

3.3 | Attempts to locate the guppy male-determining region

The LG12 region in which the guppy male-determining factor is known to be located (distal to 20 Mb) is the region that most often shows higher synonymous site diversity in males than females (Figure 2, Figure S2). However, the difference is small. Moreover, coverage is low in the region, and other high LG12 regions with diversity are clearly associated with low coverage (Figure S8). These appear to be repetitive regions, and may not reliably identify the region that includes the male-determining factor. We therefore used an approach based on associations between sex and genotypes at variable sites on LG12. The approach searches the PoolSeq data for sites (of any type, not just coding region sites) where the genotypes in males and females conform to those expected under complete sex linkage, as described in the Methods section, which also describes filtering of sites to include in the searches. Figure 5 shows results for polymorphic sites within each population, and at which variants were detected only in the population's male sample (while the females had good coverage, but were monomorphic, with XFREQ values equalling 1); the figure is based on YFREQ values above 0.9, indicating strong association with the male-determining locus (see Section 2). Although male-specific variants are concentrated in the expected region, the picture for the entire LG12 (Figure 5) does not identify a completely sex-linked region; the upper part of Figure 5 shows that at most four populations have sites satisfying this stringent Y criterion. This is consistent with the diversity results described above, and with the previously published results described in the Introduction. Figures S4 and S5A show results for male-specific variants with YFREQ = 0.7, to allow for nonfixation in the Y

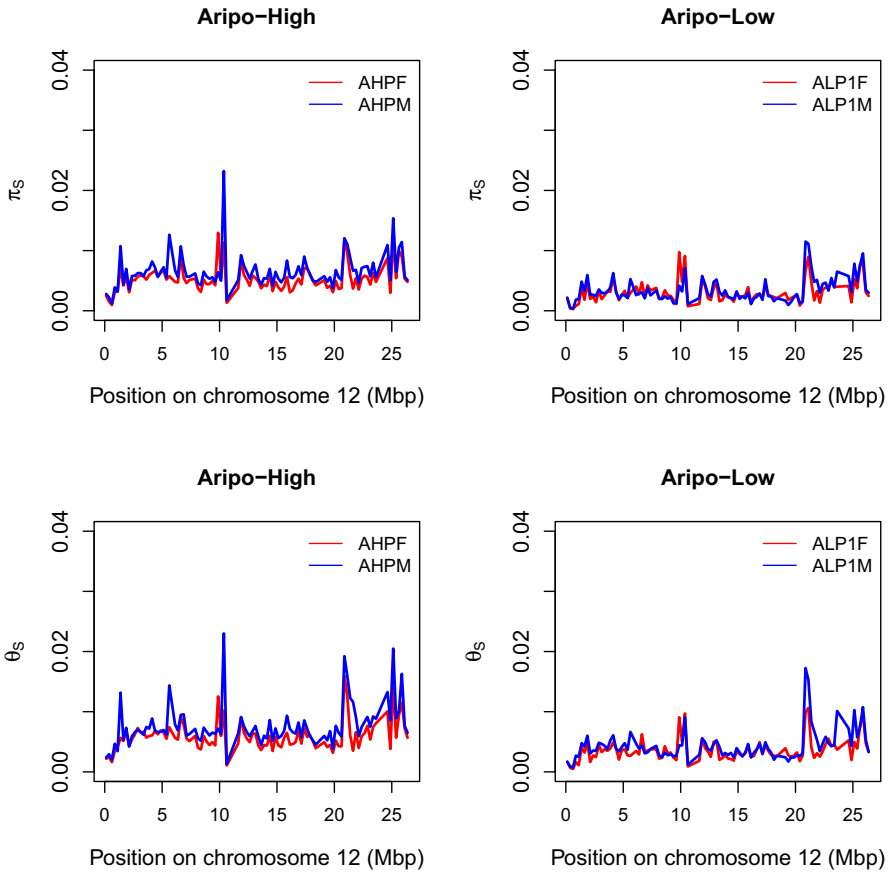

FIGURE 2 Example of diversity estimates from the Aripo river, showing the two diversity measures, both based on synonymous sites (plots for all samples analysed are in Figure S2). The values shown are mean values in 250 kbp windows, with red for female samples and blue for male ones. The green lines indicate a diversity level of 1%, to emphasize the lower diversity in a low than high predation collection site (right- vs. left-hand plots, with LP and HP in the sample names indicating high and low predation, respectively), as well its much higher θ_s than π_s value, suggesting a recent bottleneck in this LP site.

chromosome population. This does not alter the regions showing most associations with the sex-determining locus. If we include sites where one allele in our female sample matched the putatively Y-specific allele (see Section 2) the sharing of candidate fully sex-linked sites between populations did not increase (Figure S5B).

Overall, these analyses do not definitively identify either the male-determining locus or any clear fully sex-linked region in which it is located. Of the total of 7705 candidate sites with male-specific variants in at least one population sampled, 79% were detected in only one of our 12 samples. The broad signals in some samples may reflect the expected wide LD in bottlenecked populations explained above. However, some high-predation samples (including those from the Aripo and Guanapo rivers) yielded few candidate sites, and very few were detected in the Aripo LP2 sample (though their locations are consistent with the region identified in the other samples, Figure S4). Instead of a clear sex-linked region, two regions of the sex chromosome, between roughly 20.5 and 21.8 Mb, and distal to 24.5 Mb in the female assembly, are enriched in sites with apparently male-specific variants, while the signal is much weaker between 21,519,986 and 25,369,837, with no candidates shared by more than two populations. No recombinants have yet been found between markers within these three regions and the male-determining locus (Charlesworth et al., 2020), and they may all be completely sex-linked, or almost so, while the pseudoautosomal region or PAR starts more distally (slightly centromere-proximal to 25,194,513 bp, see Charlesworth et al., 2020); the regions are shown, together with information about their genetic map locations, in Figure S6, and Figure S7 shows that the two subregions with the strongest signal of sex linkage also have high repetitive content. The gap between these two regions does not reflect low coverage (Figure S8). The terminal parts of the sex chromosome may still not be correctly assembled, given their high repetitive content and presence of sequences that were unplaced in the male assembly (Figures S6 and S7), and the correct location of gap region could be distal to the two regions with male-specific SNPs.

Only 36 of the potentially male-specific sites satisfy the threshold of YFREQ ≥0.7 in at least four population samples, 29 of which are distal to 20.9 Mb. The candidate site found in the largest number of populations (at position 21,410,653) is found in only seven samples. Excluding two sites distal to the gap region suggests a candidate sex-linked region with a total size of 590,830 bp, which includes only 22 genes (Table S5). Two, cyclin and shroom3-like, are in the previously proposed candidate region (contig IV gene island), whose repetitive parts had higher coverage in males than females, and are duplicated near 24 Mb in the assembly of the sequenced male (Fraser et al., 2020). Another candidate region (Dor et al., 2019), in the more terminal LG12 region assembled distal to the gap, was also detected in this study of fish from natural populations, but no candidate genes emerged (Fraser et al., 2020).

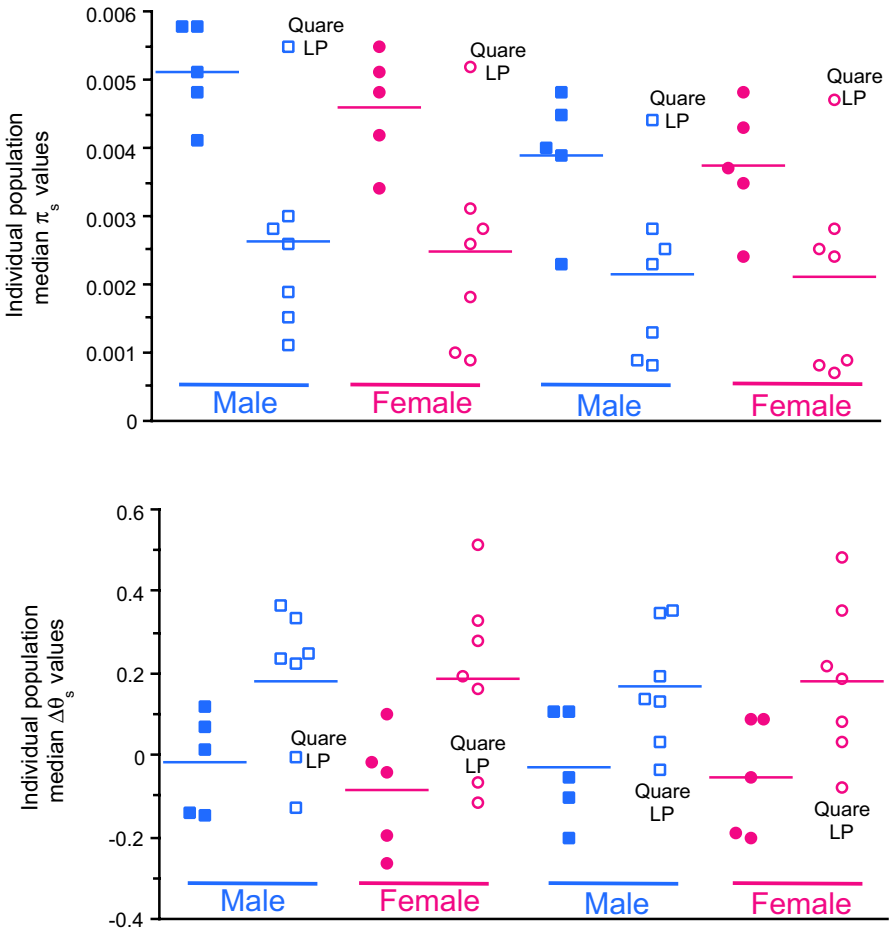

FIGURE 3 Diversity estimates in all populations sampled. Part A shows the consistently lower values of synonymous site diversity (π_s) in low-predation upstream sites (solid symbols), compared with downstream, high-predation collection sites from the same rivers (open symbols), and part B shows the higher $\Delta\theta_s$ values in the former, suggesting that upstream sites have consistently been affected by recent bottlenecks. Blue and pink symbols show male and female results, respectively. Outlier results from samples from the Quare river are labelled, with the notation HP or LP used to denote high or low predation sites, respectively, as explained for Figure 2.

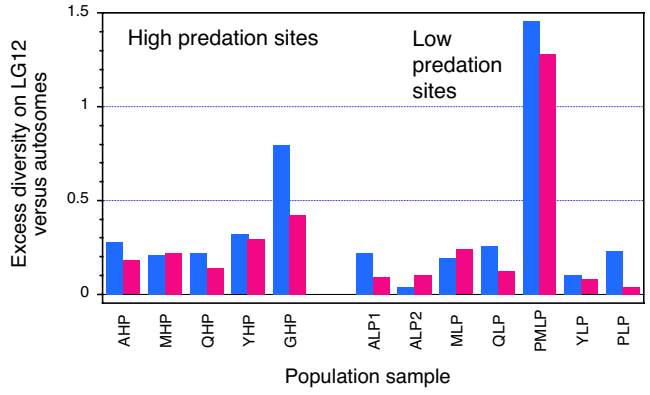

FIGURE 4 Excess of the estimated diversity on LG12, the sex chromosome, versus the autosomal estimates. The nucleotide diversity estimates are for synonymous sites (values of π_S), based on results analysed in windows (see Section 2). For each population sample shown on the *x*-axis, the plot shows values of the extent to which the LG12/autosome diversity ratio exceeds 1. Blue bars show the results for males and pink bars those for females.

4 | DISCUSSION

Our analyses of synonymous site polymorphisms indicate a greater genome-wide excess of rare variants in upstream than downstream guppy populations, indicating that at least the upstream ones are not at mutation-drift equilibrium. This will create problems for population genomic approaches aimed at detecting footprints of balancing selection (as here) or adaptive changes.

Low diversity in upstream populations is consistent with previous studies of polymorphic microsatellite or SNP markers (Barson et al., 2009; Willing et al., 2010) and very high F_{ST} values between pairs of samples from upstream populations from different rivers, though previous studies of F_{ST} either did not provide nucleotide diversity estimates (Fraser et al., 2015; Suk & Neff, 2009) or did not detect bottlenecks (Willing et al., 2010). F_{ST} estimates using the same PoolSeq data as analysed here (Yong et al., 2021; see Table S2), supported the same pattern, and a recent population genomics study using all site types inferred the occurrence of bottlenecks in several rivers (Whiting et al., 2021). Our analyses of synonymous, probably weakly selected, variants in coding sequences reveals specific loss of rare variants, as expected if upstream populations are recovering after recently losing diversity, rather than having evolved with long-term low effective population sizes. Extremely severe recent bottlenecks are excluded (they would cause almost complete loss of variability, leaving only variants at intermediate frequencies, creating a difference in $\Delta\Theta_S$ values opposite to that observed). Metapopulation dynamics, with local extinction and recolonisation events involving bottlenecks,

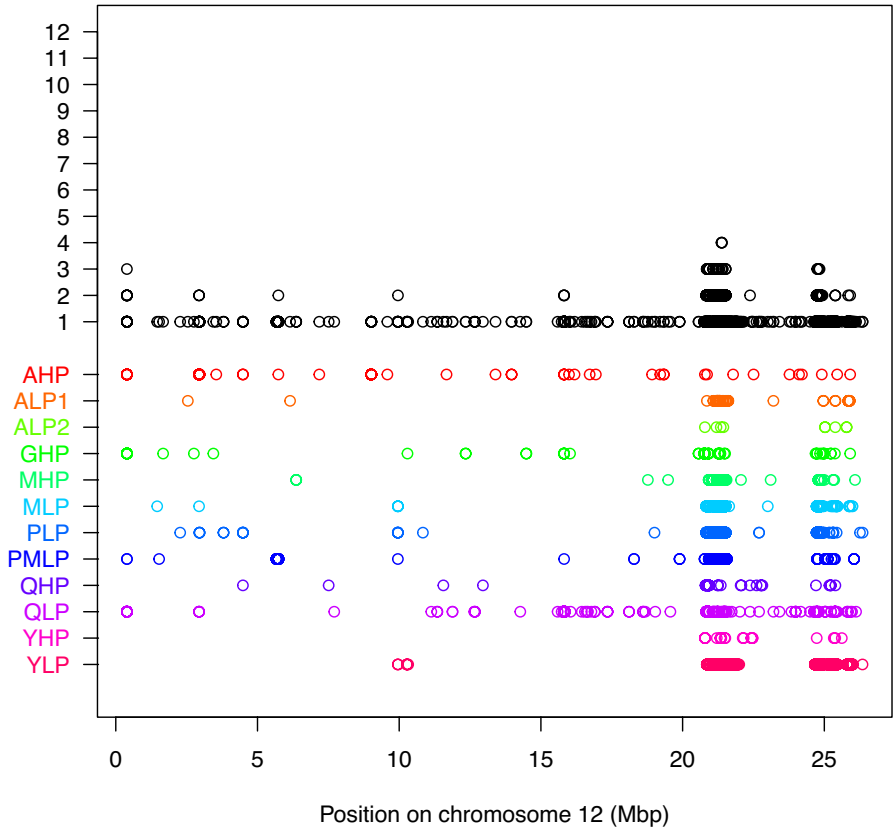

FIGURE 5 Candidate fully sex-linked sites in the entire LG12, based on PoolSeq data. SNPs at all site types (including noncoding sites) were analysed in our samples of 20 individuals of each sex from 12 natural populations. The plot shows results for sites with no variants in the 20 diploid sequences from females (XFREQ = 1 in Table 1), and the *x*-axis shows the positions of the sites in the male guppy assembly. The upper section shows the results for each such site that varied within any of the populations analysed, and the *y*-axis shows the number of populations where a site was classed as having the XY configuration, suggesting that it is a candidate for complete or near-complete sex-linkage. The results shown used the criterion that the YFREQ threshold value was 0.9, indicating that at least 18 males in the sample of 20 had the putatively Y-linked variant at the site. No site has a male-specific candidate SNP in all 12 populations. The bottom part shows the locations of candidate male-specific SNPs in each of the 12 individual population samples (in different colours, with the population codes in Table S1 shown on the *y*-axis). Signals of sex-linkage were detected in the terminal part of the chromosome in all the different populations, but are found in two regions, separated by a gap where male-specific SNPs are rare.

may contribute to the low diversity of upstream guppy populations and very high F_{ST} values between upstream sites from different drainages (Figure 3a, Tables S2 and S3).

Rare migration from downstream populations is unlikely to account for the observed consistent pattern. Unless migration rates are very low, samples of more than 10 individuals per deme in a subdivided population should not display strong departures from equilibrium (table 7.1 of Charlesworth & Charlesworth, 2010). Migration from up- to down-river localities is probably more frequent (Whiting et al., 2021), though the aberrant Quare results may reflect migration of fish from headwaters of other rivers during floods.

4.1 | Diversity differences between sex chromosomes and autosomes

Synonymous sites, under weak selective constraints reducing diversity, also consistently detect sex chromosome-specific higher diversity in males than females from natural populations (Figures 3 and 4), whereas a previous analysis of F_{ST} between guppy males and females, using all site types, did not detect clear differences between LG12 and autosomes in most populations (Fraser et al., 2020).

Diversity of LG12 sequences in females, reflecting X-linked variants, is generally likely to be less than for autosomal sequences, contrasting with our results (Figure 4). An X/A diversity ratio of 0.75 is predicted (assuming equal mutation rates for both sexes and all chromosomes) under a 1:1 sex ratio, reflecting the larger numbers of autosomes than X chromosomes in populations, and hence larger autosomal effective population size. However, this difference implies that sexual selection (reducing males' effective population size) should increase X/A ratios; this ratio can also be higher in subdivided populations, approaching 1.5 if female migration rates are much lower than male rates (Laporte & Charlesworth, 2002), though our within-population values should not be affected. Selective sweeps in X-linked genes could reduce X/A ratios (and elevate F_{ST} between males and females), but our results do not suggest recent sweeps in LG12 regions with high diversity in females or males (Figure S2).

A higher mutation rate in males than females (documented in other vertebrates in Manuel et al., 2022, but not yet for fish, to our knowledge) increases autosomal more than X-linked rates, and should also lower X/A ratios. However, synonymous and fourfold site divergence from *Xiphophorus maculatus* orthologues suggests that LG12 could have a higher neutral mutation rate than the autosomal average (Figures S9 and S10). Because the guppy Y carries alleles of X-linked genes, higher LG12 diversity within guppy populations partly reflects Y-X divergence, which contributes to these raw divergence estimates. Therefore, the effect of mutation rate differences on X/A ratios (in sequences from females) cannot be predicted without direct mutation rate estimates using parents and offspring. Differences between Y- and X-linked alleles contribute to the higher LG12 than autosomal diversity in males (LG12/A values >1 in Figure 4), and analyses based on all sites types in phased sequence data detected high Y/A diversity ratios, consistent with diversity among Y sequences also being affected by Y-X recombination (Almeida et al., 2021), as reviewed in the Introduction.

Overall, the data suggest the presence on LG12 of gene(s) under balancing selection in addition to the sex-determining locus. Even rare recombination will restrict signals of balancing selection to regions very near the male-determining locus. This is exemplified by several distantly related fish whose sex-determining factors are within recombining genome regions, the fugu (Kamiya et al., 2012), species in the genus *Seriola* (Koyama et al., 2019), and possibly also *Nothobranchius furzeri*, which also has homomorphic X and Y chromosomes (Reichwald et al., 2015). Figure 1 shows that a very low population recombination rate ($\rho < 5$) is required to account for the almost 20% higher sex chromosome than autosomal diversity observed in guppy males, in the absence of balanced polymorphisms other than that at the sex-determining locus. However, SA polymorphisms increase Y-X coalescence times at closely linked sites (Kirkpatrick & Guerrero, 2014), elevating both LG12/A and X/A ratios, as observed (Figure 4). Recombination may occur rarely in guppy males enough to allow maintenance of male coloration SA polymorphisms in any of the many hundreds of genes across most of LG12.

P. wingei appears to show LD across a larger proportion of LG12 than *P. reticulata* (Almeida et al., 2021; Darolti et al., 2019, 2020), and this may reflect recombination suppression or more male coloration factors than on the guppy Y. Samples from natural populations are not available, however, and the small sample sizes studied from this species will be even more important than in the guppy.

4.2 | Approaches for discovering completely sex-linked regions and variants

Many approaches can detect an extensive nonrecombining sex-determining region, or evolutionary stratum containing many sites with male-specific variants, but it is difficult to detect a small region, or a nonrecombining region that evolved recently, or one in which occasional recombination events still occur. Physically small male-determining loci will be difficult to locate in genomes if associations arise due to close, but incomplete, linkage across large regions, especially if linked SA polymorphisms are maintained. Low diversity populations, such as upstream guppy populations, might be expected to show the clearest signals, because associations between molecular variants and the male-determining locus should be maintained, while the rest of the genome loses variability; such differences in transposable element insertions were detected in Anole lizard populations that recently expanded after a bottleneck that reduced diversity (Bourgeois et al., 2020). The sex-linked region will thus be an outlier, compared with the rest of the genome. On the other hand, small or bottlenecked populations, with lower ρ-values, should have higher LD at equilibrium (Haddrill et al., 2005), and higher variance of LD, generating confusing footprints of sex-linkage between sequence variants that are not closely linked (Park, 2019). Some apparently male-specific variants were indeed detected proximal to 20 Mb on the guppy sex chromosome.

The guppy male-determining factor could nevertheless potentially be located by testing for LG12 regions with higher diversity. Whether the functional site is in a coding or a noncoding region, analyses of individual sites, as described here, can find variable sites with XY genotype configurations. Given a male-determining factor shared by all populations of a species, even a single base change from the sequence in females should be detectable. Associations are consistently found within a guppy sex chromosome region distal to 20 Mb. Both our results (Figure 5), and those of Almeida et al. (2021) detected somewhat stronger associations in upstream than downstream samples, as suggested above. Surprisingly, both find most of the associated SNPs in two regions, separated by a region that largely lacks such SNPs (between about 21.8 and 24.5 Mb, see Figure S6).

The approach used here may be preferable to using F_{ST} values between the sexes to detect sites associated with the sex-determining locus (Almeida et al., 2021; Bergero et al., 2019; Fraser et al., 2020). It avoids the problem that very high F_{ST} values can arise if diversity in one of the populations is low, since F_{ST} quantifies the proportion of diversity found between populations (Charlesworth, 1998; Charlesworth et al., 1997; Cruickshank & Hahn, 2014). When the sexes are treated as two populations, a recent selective sweep on the X chromosome could have reduced diversity in females, or a recent bottleneck might eliminate much X-linked diversity (leaving a high proportion of the remaining diversity due to Y-X divergence, even if that divergence is small in absolute terms).

Furthermore, if distinct Y haplotypes coexist, each with "private", haplotype-specific variants, this will reduce F_{ST} between the sexes. This situation might arise if the rarely recombining region carries polymorphic male coloration factors. Diverse Y haplotypes, associated with different male colour patterns, and supporting complete Y linkage, have been uncovered in *Poecilia* (=*Micropoecilia*) *parae* (Sandkam et al., 2021), a species with polymorphic Y-linked male

coloration (Lindholm et al., 2004). Although the male-determining factor should be shared by all males, variants in the completely linked region will have arisen independently on the different haplotypes, and many variants may not be shared with other coloration types (Figure S1). Analysing individual sites, rather than F_{ST} in windows containing multiple variable sites, partially avoids this problem, because only shared Y-linked variants that arose before the different Y haplotypes split will be detected. Reliable phasing and LD analysis can resolve such situations.

Our approach failed to find completely sex-linked variants. If the reference genome assembly of a species with male heterogamety is derived from sequencing a female, some Y-specific sequences, such as a duplication with a male-determining effect, will not be represented in the reference genome. However, use of a guppy male genome assembly also failed to reveal clear associations (Fraser et al., 2020). Regions of high repeat density may prevent Y- and X-linked sequences aligning reliably, resulting in gaps even if the male-determiner is a single nucleotide polymorphism in a gene present in both the X and Y. Our filtering to remove regions with high coverage (see Section 2) may have created gaps where repetitive sequences are abundant, potentially excluding sites with XY genotype configurations, and recent repetitive sequence insertions will create problematic haplotype-specific variants (Almeida et al., 2021). However, the absence of strong associations between about 21.8 and 24.5 Mb appears not simply to reflect this problem, as coverage values are higher than in the flanking regions with high densities of male-specific SNPs (Figure S8).

The two chromosome 12 regions in which associations with sex are commonest could reflect an assembly error, or possibly separate peaks of variation associated with the male-determining factor, and a SA polymorphism, as modelled by Kirkpatrick and Guerrero (2014), or cluster of polymorphisms. We cannot currently exclude the possibility that the different populations have different male-determining factors (Almeida et al., 2021). However, non-1:1 primary sex ratios should then arise in crosses between or within populations. Sex ratio data are scarce in guppies. To our knowledge, data are not yet available from newborn progeny or fry extracted from gravid females (which could be sexed using molecular markers located in the sex chromosome region that recombines rarely, and are mostly transmitted to male progeny of male heterozygotes). Interstrain crosses and crosses involving YY males showed that female-biased sex ratios in laboratory strains of guppies are determined primarily by Y-linked genes decreasing production or competitive ability of Y-bearing sperm, possibly reflecting accumulation of deleterious alleles, or pleiotropic effects of Y-linked alleles that increase male fitness (Farr, 1981). Finally, perhaps the genome region is arranged differently in different individuals. This could explain the different location of the signal in the Quare samples (although the signal at 20.9 Mb is weak).

Some guppy sex-linked SA factors recombine very rarely with the male-determining factor, while others show rates of up to 8%, though rates in natural populations are unclear, as many genetic studies used captive ornamental fish (Lindholm & Breden, 2002). It is currently unclear whether exchanges of coloration factors between the Y and X reflect locations in the frequently recombining region distal to the male-determining locus, or crossovers proximal to this locus. In a large region extending from the centromere, cytogenetic studies of crossover locations (Lisachov et al., 2015) and genetic mapping of molecular markers (Charlesworth et al., 2020), suggest very rare, or no, crossing over in male guppies. The polymorphic male coloration factors could be located anywhere in this region.

5 | CONCLUSIONS

Overall, our results show that there is no need to invoke complete sex linkage, or the evolution of new nonrecombining strata, to explain the pattern of differentiation between males and females across different parts of the guppy LG12. There may be no extensive (multi-gene) completely nonrecombining region corresponding to evolutionary strata of other organisms such as humans. The appearance of strata may reflect the LD that is expected across regions that rarely recombine with the male-determining locus, especially in bottlenecked populations. A male-determining locus within a 600 kb region near 21 Mb in the guppy female LG12 assembly is consistent with evidence that the male-determining factor maps distal to a marker at 21.3 Mb, while the boundary of the pseudoautosomal region is near 25 Mb (Charlesworth et al., 2020). Although the assembly of the terminal part of LG12 remains uncertain, the sex-linked region could be confined to a single SNP in or near one of the 22 genes in the region defined here (Table S5).

The localization of crossovers at one end of each chromosome arm in male guppies, may be an ancestral state, as emerging sex-specific genetic maps suggest similar patterns in other fish (Sardell & Kirkpatrick, 2020), and there is currently no evidence for changed recombination rates in male guppies. Therefore, rather than selecting for suppressed recombination between loci with SA balanced polymorphisms and the male-determining locus, low recombination across most of LG12 (producing close linkage with the male-determining factor), may have favoured establishment of such polymorphisms. The resulting associations of variants across the XY pair (Kirkpatrick & Guerrero, 2014) would resemble young evolutionary strata.

LD across the sex chromosome could reflect other SA polymorphisms not just male coloration factors. Guppy populations show several other sexual dimorphisms, including smaller size of males than females and behavioural differences (Olendorf et al., 2006). However, it is currently unclear whether the differentiation detected across the guppy sex chromosomes implies the presence of multiple SA factors. Theoretical modelling shows that this requires strong selection at loci that are tightly linked, but distant from the sex determining region, and the loci remain polymorphic only if selection favours heterozygotes (Otto, 2019). Nucleotide diversity depends on recent demographic changes, not just on the recombination rates and effective population sizes (i.e., ρ-values in Figure 1) under which populations are evolving. Upstream populations have undergone such changes, and also evolutionary changes since these sites were colonized, as reviewed by

Magurran (2005), and selective sweeps may have affected diversity of some genome regions. Therefore, even if the low recombination rates across much of the guppy LG12 could be estimated, it may be difficult to fit diversity results to theoretical equations to test whether LD with weakly selected synonymous variants is so high that SA polymorphisms must be invoked to account for it.

AUTHOR CONTRIBUTIONS

Deborah Charlesworth designed research. Lengxob Yong, Alastair Wilson, and Darren P. Croft performed research. Suo Qiu, Lengxob Yong and Chay Graham analysed the data. Deborah Charlesworth, Sue Qiu, Alastair Wilson and Darren P. Croft wrote the manuscript.

ACKNOWLEDGEMENTS

This project was supported by ERC grant number 695225 (GUPPYSEX). We acknowledge help from the Exeter Advanced Computing Resources (Athena), and thank I.W. Ramnarine and R. Mahabir (University of the West Indies, Trinidad and Tobago) for assistance with the field collection, and Mateusz Konczal (Adam Mickiewicz University, Poznań, Poland) for helpful suggestions about analyses.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Genomic data (raw fastq files) have been made available under ENA Project Accession PRJEB45804. Benefits Generated: Collaborating scientists from the University of the West Indies, Trinidad, R. Mahabir and R. Heathcote assisted with the collection of the natural population samples and Trinidad-UK shipment of guppies, respectively, as acknowledged in Yong et al. (2021).

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Lengxob Yong https://orcid.org/0000-0002-7459-4863
Alastair Wilson https://orcid.org/0000-0002-5045-2051
Deborah Charlesworth https://orcid.org/0000-0002-3939-9122

REFERENCES

Almeida, P., Sandkam, A. A., Morris, J., Darolti, I., Breden, F., & Mank, J. (2021). Divergence and remarkable diversity of the Y chromosome in guppies. *Molecular Biology and Evolution*, *38*(2), 619–633. https://doi.org/10.1093/molbev/msaa257

Barson, N., Cable, J., van Oosterhout, C., Joyce, D., Cummings, S., Blais, J., Barson, N., Ramnarine, I., Mohammed, R., & Persad, N. (2009). Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *Journal of Evolutionary Biology*, *22*, 485–497. https://doi.org/10.1111/j.1420-9101.2008.01675.x

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*(14), 6924–6931. https://doi.org/10.1073/pnas.1818486116

Bourgeois, Y., Ruggiero, R. P., Hariyani, I., & Boissinot, S. (2020). Disentangling the determinants of transposable elements dynamics in vertebrate genomes using empirical evidences and simulations. *PLoS Genetics*, *16*(10), e1009082. https://doi.org/10.1371/journal.pgen.1009082

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*, 538–543.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided inbreeding and outcrossing populations. *Genetical Research*, *70*, 155–174.

Charlesworth, D. (2019). Young sex chromosomes in plants and animals. *New Phytologist*, *224*(3), 1095–1107. https://doi.org/10.1111/nph.16002

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

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

Charlesworth, D., & Charlesworth, B. (2010). *Elements of evolutionary genetics*. Roberts and Company.

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, *508*, 488–493. https://doi.org/10.1038/nature13151

Cruickshank, T., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Dagilis, A. J., Sardell, J., Josephson, M., Su, Y., Kirkpatrick, M., & Peichel, C. L. (2022). Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*, 20210205. https://doi.org/10.1098/rstb.2021.0205

Darolti, I., Wright, A., & Mank, J. (2020). Guppy Y chromosome integrity maintained by incomplete recombination suppression. *Genome Biology and Evolution*, *12*(6), 965–977. https://doi.org/10.1093/gbe/evaa099

Darolti, I., Wright, A., Sandkam, B., Morris, J., Bloch, N., Farré, M., Fuller, R., Bourne, G., Larkin, D., Breden, F., & Mank, J. E. (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. https://doi.org/10.1073/pnas.1905298116

Dor, L., Shirak, A., Kohn, Y., Gur, T., Welle, J., Zilberg, Z., Seroussi, E., & Ron, M. (2019). Mapping of the sex determining region on linkage group 12 of guppy (*Poecilia reticulata*). *G3 (Bethesda)*, *9*, 3867–3875. https://doi.org/10.1534/g3.119.400656

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

Farr, J. (1981). Biased sex ratios in laboratory strains of guppies, *Poecilia reticulata*. *Heredity*, *47*(2), 237–248.

Fraser, B., Hughes, K., Tosh, D., & Rodd, F. (2013). The role of learning by a predator, *Rivulus hartii*, in the rare-morph survival advantage in guppies. *Journal of Evolutionary Biology*, *26*(12), 2597–2605. https://doi.org/10.1111/jeb.12251

Fraser, B. A., Künstner, A., Reznick, D. N., Dreyer, C., & Weigel, D. (2015). Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). *Molecular Ecology*, *24*, 389–408. https://doi.org/10.1111/mec.13022

Fraser, B. A., Whiting, J. R., Paris, J. R., Weadick, C. J., Parsons, P. J., Charlesworth, D., Bergero, R., Bemm, F., Hoffmann, M., Kottler, V., Liu, C., Dreyer, C., & Weigel, D. (2020). Improved reference genome uncovers novel sex-linked regions in the guppy (*Poecilia reticulata*). *Genome Biology and Evolution*, *12*, 1789–1805. https://doi.org/10.1093/gbe/evaa187

Gordon, S. P., López-Sepulcre, A., & Reznick, D. N. (2012). Predation-associated differences in sex-linkage of wild guppy coloration. *Evolution*, *66*(3), 912–918. https://doi.org/10.1111/j.1558-5646.2011.01495.x

Gordon, S. P., López-Sepulcre, A., Rumbo, D., & Reznick, D. N. (2017). Rapid changes in the sex linkage of male coloration in introduced guppy populations. *American Naturalist*, *189*(2), 196–200. https://doi.org/10.1086/689864

Haddrill, P. R., Thornton, K. R., Charlesworth, B., & Andolfatto, P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research*, *15*(6), 790–799. https://doi.org/10.1101/gr.3541005

Haskins, C., Haskins, E. F., McLaughlin, J., & Hewitt, R. E. (1961). Polymorphisms and population structure in *Lebistes reticulatus*, an ecological study. In W. F. Blair (Ed.), *Vertebrate speciation* (pp. 320–395). University of Texas Press.

Hivert, V., Leblois, R., Petit, E., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics*, *210*(1), 315–330. https://doi.org/10.1534/genetics.118.300900

Houde, A. E. (1992). Sex-linked heritability of a sexually selected character in a natural population of *Poecilia reticulata* (Pisces: Poeciliidae) (guppies). *Heredity*, *69*, 229–235. https://doi.org/10.1038/hdy.1992.120

Hughes, K., Houde, A., Price, A., & Rodd, F. (2013). Mating advantage for rare males in wild guppy populations. *Nature*, *503*, 108–110. https://doi.org/10.1038/nature12717

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, M., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., & Kikuchi, K. (2012). A trans-species missense SNP in amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, *8*, e1002798. https://doi.org/10.1371/journal.pgen.1002798

Kirkpatrick, M., Guerrero, R., & Scarpino, S. (2010). Patterns of neutral genetic variation on recombining sex chromosomes. *Genetics*, *184*(4), 1141–1152. https://doi.org/10.1534/genetics.109.113555

Kirkpatrick, M., & Guerrero, R. (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics*, *197*, 531–541. https://doi.org/10.1534/genetics.113.156026

Kofler, R., Orozco-terWengel, P., Maio, N. D., Pandey, R., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, *6*(1), e15925. https://doi.org/10.1371/journal.pone.0015925

Koyama, T., Nakamoto, M., Yamashita, R., Yamashita, T., Sasaki, K., Kuruma, Y., Mizuno, N., Suzuki, M., & Okada, Y. (2019). A SNP in a steroidogenic enzyme is associated with phenotypic sex in Seriola fishes. *Current Biology*, *29*, 1901–1909. https://doi.org/10.1016/j.cub.2019.04.069

Lahn, B. T., & Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science*, *286*, 964–967. https://doi.org/10.1126/science.286.5441.964

Laporte, V., & Charlesworth, B. (2002). Effective population size and population subdivision in demographically structured populations. *Genetics*, *162*, 501–519.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindholm, A., & Breden, F. (2002). Sex chromosomes and sexual selection in Poeciliid fishes. *American Naturalist*, *160*, S214–S224. https://doi.org/10.1086/342898

Lindholm, A., Brooks, R., & Breden, F. (2004). Extreme polymorphism in a Y-linked sexually selected trait. *Heredity*, *95*, 156–162. https://doi.org/10.1038/sj.hdy.6800386

Lisachov, A., Zadesenets, K., Rubtsov, N., & Borodin, P. (2015). Sex chromosome synapsis and recombination in male guppies. *Zebrafish*, *12*(2), 174–180. https://doi.org/10.1089/zeb.2014.1000

Magurran, A. E. (2005). *Evolutionary ecology: The Trinidadian guppy*. Oxford University Press.

Manuel, M., Wu, F. L., & Przeworski, M. (2022). A paternal bias in germline mutation is widespread across amniotes and can arise independently of cell divisions. *bioRxiv*. https://doi.org/10.1101/2022.02.07.479417

Nanda, I., Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., & Schartl, M. (2014). Sex chromosome polymorphism in guppies. *Chromosoma*, *123*(4), 373–383. https://doi.org/10.1007/s00412-014-0455-z

Nelson, C., Moncla, L., & Hughes, A. (2015). SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics*, *31*(22), 3709–3711. https://doi.org/10.1093/bioinformatics/btv449

Olendorf, R., Rodd, F., Punzalan, D., Houde, A., Reznick, D., & Hughes, K. (2006). Frequency-dependent survival in natural guppy populations. *Nature*, *441*, 633–636. https://doi.org/10.1038/nature04646

Otto, S. (2019). Evolutionary potential for genomic islands of sexual divergence on recombining sex chromosomes. *New Phytologist*, *224*(3), 1241–1251. https://doi.org/10.1111/nph.16083

Papadopulos, A. S. T., Chester, M., Ridout, K., & Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(42), 13021–13026. https://doi.org/10.1073/pnas.1508454112

Park, L. (2019). Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. *Scientific Reports*, *9*, 11380. https://doi.org/10.1038/s41598-019-47832-y

Reichwald, K., Lauber, C., Nanda, I., Kirschner, J., Hartmann, N., Schories, S., Gausmann, U., Taudien, S., Schilhabel, M., Szafranski, K., Glöckner, G., Schmid, M., Cellerino, A., Schartl, M., Englert, C., & Platzer, M. (2009). High tandem repeat content in the genome of the short-lived annual fish Nothobranchius furzeri: a new vertebrate model for aging research. *Genome Biology*, *10*, R16. https://doi.org/10.1186/gb-2009-10-2-r16

Reichwald, K., Petzold, A., Koch, P., Downie, B., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., Sahm, A., Szafranski, K., Taudien, S., Groth, M., Arisi, I., Weise, A., Bhatt, S. S., Sharma, V., Kraus, J. M., … Platzer, M. (2015). Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell*, *163*, 1527–1538. https://doi.org/10.1016/j.cell.2015.10.071

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex-chromosomes. *Evolution*, *41*, 911–914.

Sandkam, B. A., Almeida, P., Darolti, I., Furman, B., van der Bijl, W., Morris, J., Bourne, G., Breden, F., & Mank, J. (2021). Extreme Y chromosome polymorphism corresponds to five male reproductive morphs. *Nature Ecology & Evolution*, *5*, 939–948. https://doi.org/10.1038/s41559-021-01452-w

Sardell, J., & Kirkpatrick, M. (2020). Sex differences in the recombination landscape. *American Naturalist*, *195*, 361–379. https://doi.org/10.1086/704943

Schmidt, J. (1920). The genetic behavlour of a secondary sexual character. *Comptes-rendus des travails du Laborotoire Carlsberg*, *14*, 8.

Stoler, N., & Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics*, *3*(1), lqab019. https://doi.org/10.1093/nargab/lqab019

Suk, H., & Neff, B. D. (2009). Microsatellite genetic differentiation among populations of the Trinidadian guppy. *Heredity*, *102*, 425–434. https://doi.org/10.1038/hdy.2009.7

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis. *Genetics*, *123*, 585–595.

Traut, W., & Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: Zebrafish, platyfish and guppy. *Chromosome Research*, *9*(8), 659–672. https://doi.org/10.1023/A:1012956324417

Tripathi, N., Hoffmann, M., Weigel, D., & Dreyer, C. (2009). Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (*Poecilia reticulata*). *Genetics*, *182*, 365–374. https://doi.org/10.1534/genetics.108.098541

Whiting, J. R., Paris, J. R., Zee, M. J., Parsons, P. J., Weigel, D., & Fraser, B. A. (2021). Drainage-structuring of ancestral variation and a common functional pathway shape limited genomic convergence in natural high- and low-predation guppies. *PLoS Genetics*, *17*, e1009566. https://doi.org/10.1371/journal.pgen.1009566

Willing, E.-M., Bentzen, P., van Oosterhout, C., Hoffmann, M., Cable, J., Breden, F., Weigel, D., & Dreyer, C. (2010). Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology*, *19*, 968–984. https://doi.org/10.1111/j.1365-294X.2010.04528.x

Winge, O. (1922). A peculiar mode of inheritance and its cytological explanation. *Journal of Genetics*, *12*, 137–144.

Wright, A., Darolti, I., Bloch, N., Oostra, V., Sandkam, B., Buechel, S., Kolm, N., Breden, F., Vicoso, B., & Mank, J. (2017). Convergent recombination suppression suggests a role of sexual conflict in guppy sex chromosome formation. *Nature Communications*, *8*, 14251. https://doi.org/10.1038/ncomms14251

Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., Zhang, G., Blom, P., Christidis, L., Prost, S., Irestedt, M., & Zhou, Q. (2019). Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nature Ecology and Evolution*, *3*, 834–844. https://doi.org/10.1038/s41559-019-0850-1

Yong, L., Croft, D. P., Troscianko, J., Ramnarine, I. W., & Wilson, A. (2021). Sensory-based quantification of male colour patterns in Trinidadian guppies reveals no support for parallel phenotypic evolution in multivariate trait space. *Molecular Ecology*, *31*, 1337–1357. https://doi.org/10.1111/mec.16039

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Qiu, S., Yong, L., Wilson, A., Croft, D. P., Graham, C., & Charlesworth, D. (2022). Partial sex linkage and linkage disequilibrium on the guppy sex chromosome. *Molecular Ecology*, *00*, 1–14. https://doi.org/10.1111/mec.16674