Has recombination changed during the recent evolution of the guppy Y chromosome?

Charlesworth, Deborah; Qiu, Suo; Bergero, Roberta; Gardner, Jim; Keegan, Karen; Yong, Lengxob; Hastings, Abigail; Konczal, Mateusz

Published in:
Genetics

DOI:
10.1093/genetics/iyad198

First published: 13/11/2023

Document Version
Peer reviewed version

Has recombination changed during the recent evolution of the guppy Y chromosome?

Deborah Charlesworth (ORCID iD 0000-0002-3939-9122), Suo Qiu, Roberta Bergero [1], Jim Gardner, Karen Keegan, Lengxob Yong [2,3], Abigail Hastings, Mateusz Konczal [4]

Institute of Ecology and Evolution, School of Biological Sciences,
University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3LF, U.K.

[1] Scottish Rural Agricultural College, Peter Wilson Building, King's Buildings, W Mains Rd, Edinburgh EH9 3JG

[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] Evolutionary Biology Group, Faculty of Biology, Adam Mickiewicz University; 60-614 Poznań, Poland

Corresponding author: Deborah.Charlesworth@ed.ac.uk

Keywords

Partial sex linkage, pseudo-autosomal region, sexually antagonistic polymorphism, sex reversal

Running head: Guppy sex chromosome genetic map

Abstract

Genome sequencing and genetic mapping of molecular markers have demonstrated nearly complete Y-linkage across much of the guppy (*Poecilia reticulata*) XY chromosome pair. Predominant Y-linkage of factors controlling visible male-specific coloration traits also suggested that these polymorphisms are sexually antagonistic (SA). However, occasional exchanges with the X are detected, and recombination patterns also appear to differ between

natural guppy populations, suggesting ongoing evolution of recombination suppression under selection created by partially sex-linked SA polymorphisms. We used molecular markers to directly estimate genetic maps in sires from four guppy populations. The maps are very similar, suggesting that their crossover patterns have not recently changed. Our maps are consistent with population genomic results showing that variants within the terminal 5 Mb of the 26.5 Mb sex chromosome, chromosome 12, are most clearly associated with the maleness factor, albeit incompletely. We also confirmed occasional crossovers proximal to the male-determining region, defining a second, rarely recombining, pseudo-autosomal region, PAR2. This fish species may therefore have no completely male-specific region (MSY) more extensive than the male-determining factor. The positions of the few crossover events suggest a location for the male-determining factor within a physically small repetitive region. A sex-reversed XX male had few crossovers in PAR2, suggesting that this region's low crossover rate depends on the phenotypic, not the genetic, sex. Thus, rare individuals whose phenotypic and genetic sexes differ, and/or occasional PAR2 crossovers in males can explain the failure to detect fully Y-linked variants.

Introduction

Guppy (*Poecilia reticulata*) populations have been important for studies relating to adaptive evolution and the evolution of sex chromosomes, especially the rarity or absence of recombination between the Y- and X-linked regions. More than 100 years ago, the Y chromosome of this sexually dimorphic fish was shown to carry a male coloration factor (Schmidt 1920), and more such factors were later identified as polymorphisms in natural populations (reviewed in Winge 1927). Although such coloration is associated with higher predation rates, frequency-dependent advantages in males (rare coloration phenotypes gain matings and survive better than ones at higher frequencies) help these factors to be maintained at intermediate frequencies in nature (Haskins and Haskins 1951; recent studies confirming these conclusions are reviewed in Potter *et al.* 2023). Y-linkage suggests that the mutations involved were sexually antagonistic (SA) when they first arose: the higher predation rates probably reduce female survival, creating a classical sexually antagonistic situation with an advantage in males and disadvantage in females.

If recombination occasionally occurs (so that SA coloration factors were initially partially sex linked), selection favors closer linkage (Charlesworth and Charlesworth 1980; Rice 1987; Wright *et al.* 2017). This could have led to the evolution of a region completely linked with the guppy male-determining factor (no recombination), forming a male-specific region, or MSY (see Figure 1A) (reviewed in Wright *et al.* 2017). This possibility was supported by evidence that guppy populations under differing predation pressure might also differ in the Y-linkage of coloration factors. Here, we examine the previous, mostly indirect evidence for inter-population differences in Y-linkage, and then describe the specific differences that are expected between guppy populations and can be tested by our direct genetic mapping.

The involvement of SA polymorphisms in guppies was based on the following evidence for predominant Y-linked control of the presence of individual male coloration traits. First, Winge's (1927) genetic inferences were based on father-to-son transmission (44 sires and 749 sons) of natural male coloration traits found in rivers in Trinidad. Haskins *et al.* (1961) confirmed Y-linkage using experiments based on the fact that coloration factors can be detected in females by testosterone treatment. In 537 treated female progeny of wild males, Haskins *et al.* (1961) estimated (in their Table 12) that 75% of the patterns showed Y-linked transmission (female progeny did not show their sires' traits), while 10% were autosomal. Given that the guppy has 23 chromosomes, and that the sex chromosome, chromosome 12, carries less than 5% of the species' genes (Künstner *et al.* 2017), Y chromosomal control of these patterns appear to be strongly over-represented. Winge (1934) also showed that most male coloration traits are Y-linked, though sex linkage is partial. He found one recombinant between the factors controlling the red and the black elements of the *maculatus* pattern among 3,800 progeny studied. Lindholm and Breden (2002) reviewed all available studies, including those in domesticated guppies. Several studies confirmed that Y-X recombination occurs, albeit at low rates. In one multi-factor linkage map, the variegated tail factor (*Var*) is on one side of the sex-determining locus, and genes for 10 other visible traits are within < 15 cM on the other side, suggesting the presence of two pseudo-autosomal regions (Khoo *et al.* 1999). This early evidence that recombination is not completely suppressed in the guppy is confirmed by the finding that its Y

chromosome has not undergone genetic degeneration (Bergero *et al.* 2019; Fraser *et al.* 2020; Almeida *et al.* 2021; Charlesworth *et al.* 2021).

The properties just described (predominant Y-linkage, and male-specific expression) both support the view that, the mutations that created the coloration factors were sexually antagonistic and reduced fitness in females. Mutations expressed only in males could spread in populations even if they were autosomal, so they should not show the observed concentration on the Y chromosome. Moreover, both Y-linked and autosomal sexually antagonistic factors may subsequently evolve sex biased expression.

Y-linkage of individual male coloration traits has, however, recently been questioned by a quantitative trait analysis of a number of male coloration traits (Morris *et al.* 2020). This requires evaluation, as it questions the reasoning just outlined supporting SA selection acting on coloration factors. However, excluding traits shared by almost all sires, which are uninformative for genetic study, and tail traits (which are often not Y-linked), analyses of individual orange and black spots in a captive population derived from a high-predation site of the Quare river in Trinidad found a major Mendelian factor for the orange spot (OI) trait, which showed Y-linkage (Morris *et al.* 2020). Importantly, this study also detected large grandpaternal effects for composite orange and black spot area phenotypes (versus mostly negative grandmaternal effects), suggesting a strong paternal transmission component, especially given that non-variable traits were included. This is consistent with other studies reviewed by Morris et al. (2020), including evidence that modifiers are involved (Haskins and Haskins 1951; Tripathi *et al.* 2009), and with analyses using guppies collected in Australia, which estimated that about half of the genetic effects represent Y-linkage (Postma *et al.* 2011). Overall, therefore, large Y-linked effects appear to be convincingly demonstrated.

Large Y-linked effects on male coloration phenotypes, together with the evidence supporting rare Y-X recombination suggested that partially sex-linked factors controlling SA male coloration traits might have created selection for suppressed recombination, creating new sex chromosome strata in populations in which predation pressure most strongly disadvantages females expressing coloration traits (reviewed in Wright *et al.* 2017). Several studies have reported evidence for inter-population differences in Y-X recombination in Trinidadian guppies,

suggesting ongoing evolution of crossover patterns and/or rates. First, testosterone treatment has consistently revealed higher frequencies of female carriers of several coloration factors in upstream populations, with low predation rates, than downstream, with high predation rates (page 374 of Haskins *et al.* 1961). Figure S1 summarizes similar results from further experiments using treated females from five rivers (Gordon *et al.* 2012). This is indirect evidence that recombination rates in male meiosis could differ between high- and low-predation populations. However, other explanations are not excluded. For example, the populations may differ in the expression of coloration traits in testosterone treated females, and/or in autosomal factors suppressing expression in females carrying coloration factors (see the Discussion section).

For the widely distributed Sb coloration polymorphism, whose frequency is generally below 20% (Table 11 of Haskins *et al.* 1961), indirect tests also suggested a recombination rate difference (Figure S1), and a difference was also demonstrated directly. Thirty-three males from a downstream population in the Aripo river (with high predation) transmitted the factor almost exclusively to male progeny, demonstrating Y-linkage. However, nearly all the testosterone treated female progeny from four of 19 Sb males from a low-predation site expressed the Sb phenotype, demonstrating that these males' X chromosomes carried the factor. Seven other males carried the factor on both the X and Y (all male progeny and treated female progeny had the Sb phenotype; Table 14 in Haskins *et al.* 1961). In these families, only a single potentially recombinant individual was found among 1,024 progeny of downstream males, versus 4 among 387 progeny of upstream males. Although both proportions of recombinants are low, there is a 13-fold difference (P = 0.0071 in a two tailed Fisher's exact test), and *Sb* is therefore almost completely Y-linked only in the downstream population.

Population genomic analyses also suggested between-population recombination differences (Wright *et al.* 2017); a sub-terminal region appeared likely to be a completely sex-linked region that became established after the guppy male-determining factor (or "M factor") evolved within the region (Figure 1). This would represent an oldest "evolutionary stratum", like the strata in the human XY pair detected through Y-specific variants accumulated in completely sex-linked sequences other than the male-determining factor itself (Lahn and Page

1999); such completely sex-linked regions are often termed male-specific (or MSYs; Figure 1A). Wright *et al.* (2017) also suggested that upstream guppy populations later evolved new strata by stopping recombination across approximately 10 Mb centromere-proximal to the proposed old stratum (Figure 1A). However, analyses of larger samples suggest that sex-associated SNPs (indicating linkage disequilibrium with the maleness factor) are rare before 20 Mb in the assembly (Almeida *et al.* 2021; Qiu *et al.* 2022), which is supported by evidence for Y-X recombination using phylogenetic analyses (Kirkpatrick *et al.* 2022). Limitation of linkage disequilibrium (LD) to the distal part of chromosome 12 confirms that crossovers occasionally occur in the region centromere-proximal to the M factor.

These findings are consistent with cytogenetic studies in male meiosis (Lisachov *et al.* 2015) and results of genetic mapping of SNPs. The terminal region is a highly recombining PAR1 (partially sex-linked), while much of the rest of the chromosome recombines rarely (Bergero *et al.* 2019; Charlesworth *et al.* 2020b). Crossover localization to the centromere-distal telomeric ends of the chromosomes is not restricted to the sex chromosome pair and probably reflects an ancestral pattern (Charlesworth *et al.* 2020b). The distal PAR1 part of the XY pair occupies around 4% of the X chromosome assembly (Künstner *et al.* 2017). As in other species with strong crossover localization in physically small PARs, such as the human PAR1 (Rouyer *et al.* 1986), and the mouse and collared flycatcher PARs (Marais and Galtier 2003; Smeds *et al.* 2014), the guppy PAR1 has very high crossover rates. The remaining "XY region" must include the male-determining factor and possibly also a wider MSY region (see Figure 1A), or the M factor could be a small region within the region of rare recombination in male meiosis, which we refer to as PAR2.

It is important to note that recombination suppression is not the only possible explanation for predominant Y-linkage of male coloration factors. Importantly, establishment of a SA polymorphism depends on the recombination rate between the gene involved and the male-determining locus (Jordan and Charlesworth 2012). This is an example of the general result that the fate of a mutation that is favored only in combination with one allele at a previously established polymorphic locus depends on its linkage to that locus (Fisher 1931). Such a "selective sieve" or "linkage constraint" can account for linked polymorphisms involved

in Batesian mimicry (Charlesworth and Charlesworth 1975; Turner 1977), and sex differences (Charlesworth and Charlesworth 1978); less closely linked mutations with identical phenotypic effects may either fail to spread, or spread to fixation. Male coloration mutations that become polymorphic within guppy populations should therefore cluster within a genome region closely linked to the male-determining locus, in LD with it. Predominant Y-linkage therefore does not require any change in recombination, although such two-locus polymorphisms favour closer linkage. Regions of the Y chromosome that recombine rarely with the X are especially likely to evolve such situations.

It therefore remains unclear whether closer linkage to the guppy male-determining factor has evolved, or whether the present recombination pattern pre-dates establishment of polymorphic coloration factors. If recombination does differ between different guppy populations, and evolution of crossover patterns is ongoing, the changes must probably affect the distal region of the XY pair, and two types of change are possible (Figure 1): there could be a difference in the recombination rates within PAR1 (Figure 1B), or the PAR1 boundary might differ (Figure 1C).

Here we directly test for differences in the genetic maps from males from upstream and downstream populations from the Aripo and Quare rivers, using molecular markers. We used the guppy female and male assemblies (Künstner *et al.* 2017; Fraser *et al.* 2020) to relate male meiotic genetic map locations to the physical positions of markers, especially in the distal region of the sex chromosome. As described above, the only direct recombination rate estimates are very low in males from both high- and low-predation populations (possibly because, although the *Sb* locus appears to be within PAR1, is very close to the boundary with the XY region); the difference between the populations studies is therefore small. Mapping of molecular variants can include markers at larger genetic distances from this boundary, which should make recombination rate differences much more detectable.

Methods

Families used for genetic mapping

Table S1 lists the sources of the parents, all of which were captured in the natural populations indicated in the table, and the numbers of parents and progeny used for studying recombination. Detailed maps were estimated for two Aripo river families, and two families from the Quare river, in each case from one high- and one low-predation site (indicated by H or HP in the family name for the former, and by LP for the latter, see Table S1). Crosses were made between the parents after these were transported to the University of Exeter, Penryn, except for the family from the Aripo high-predation site, whose parents were from a captive population that had been maintained for several years since being collected (Bergero *et al.* 2019; Charlesworth *et al.* 2020a). All the families except this one consisted of several full sibships with different parental individuals. In sibships with multiple possible parents, parentage of progeny was inferred using genotypes at multiple microsatellite markers and confirmed using SNPs in the high-throughput genotyping data (see below).

Genetic markers

SNPs for genetic mapping were genotyped in high-throughput experiments performed by LGCG, https://www.lgcgroup.com/ (Charlesworth *et al.* 2020a). To maximise the chance of getting markers in single-copy sequences, the targets were selected in coding sequences annotated in the guppy female genome assembly (Künstner *et al.* 2017), using NCBI Poecilia reticulata 1.0: https://www.ncbi.nlm.nih.gov/assembly/GCF_000633615.1/. The mapping results were obtained in three separate experiments, as samples from the families became available. For our first genetic mapping experiment, Experiment 2, a set of about 500 target coding sequences were selected from the sex chromosome pair, and smaller numbers from each of the 23 guppy autosomes. Experiment 3 used a larger number of SNPs from the sex chromosome pair, and Experiment 4 mapped the same set of SNPs from the sex chromosome pair, plus large numbers from the autosomal chromosome 16. Because the male genome assembly is not annotated, the male assembly locations were found by BLAST searches, as the male and female assemblies are not perfectly syntenic. This identified most, but not all, sequences targeted (See Supplementary Tables 3C and D). For sex-linked targets that were not in the chromosome 12 male assembly, microsatellite markers were designed for mapping in Experiment 2 (Supplementary Table 2B).

Data pre-processing by LGCG included demultiplexing of all libraries for each sequencing lane using the Illumina bcl2fastqv2.20 software [Illumina. bcl2fastq2 Conversion Software. URL: https://support.illumina.com/sequencing/sequencing software/bcl2fastq-conversion-software.html] (folder 'RAW' subfolders), allowing at most 2 mismatches or Ns in the barcode read when the barcode distances between all libraries on the lane allowed for it. The sequencing adapter remnants were then clipped from all reads (folder "AdapterClipped") and reads with final lengths < 20 bases were discarded before merging forward and reverse reads using BBMerge v34.48 citebbtools (folder "Combined"); the consensus sequence of combinable fragments are named "*joined-SR*", and sequences with read pairs that could not be combined were stored in files named "*R1*" and "*R2*". The remaining (combined) reads were quality cropped, setting the quality of the first 40 bps (corresponding to the synthetic oligonucleotide probe) to 0 to avoid biased genotyping in the downstream analysis, and the resulting sequences are in the "QualityTrimmed" folder. FastQC reports were created for all FASTQ files (Andrews 2010). Finally, read counts for all samples were generated. This involved initial alignment of subsampled quality trimmed reads against the reference sequence using BWA-MEM v0.7.12, http://bio-bwa.sourceforge.net/ (Li 2013), followed by variant detection and genotyping of all samples using Freebayes v1.2.0, http://github.com/ekg/freebayes (Garrison and Marth 2012), with ploidy set to 2, since the guppy Y chromosome is not genetically degenerated. Genotypes were filtered to a minimum coverage of 8 reads.

Microsatellite sequences were found in the guppy female and male genome assemblies (Künstner *et al.* 2017; Fraser *et al.* 2020) using the PERF software v0.4.5 (Avvaru *et al.* 2018). Excel files listing microsatellites for all guppy chromosomes, based on both the male and female assemblies, including the PCR primer sequences for amplifying them, and files with the marker locations and the genetic mapping results have been deposited in Figshare (https://figshare.com/account/collections/6644282).

Genetic mapping

The high-throughput genotyping experiments produced counts of each allele inferred to be present in each individual from the chromosomes targeted. Duplicate DNA samples of some

individuals were included, to assess the reliability of the results. These files, with the marker locations, and the genetic mapping results are also available in Figshare. The marker results were analysed using the LepMap3 software package, which is designed for mapping data from families consisting of half-sibships with shared parents (Rastas 2017). The genotypes were analysed as autosomal markers, because the guppy Y chromosome is not genetically degenerated, and appears to carry all genes present on the X (Bergero *et al.* 2019; Fraser *et al.* 2020; Almeida *et al.* 2021; Charlesworth *et al.* 2021). Before estimating genetic maps, the total allele counts for all individuals in a family to be analysed were filtered to exclude sequences with low coverage (the thresholds are explained in the headers in Supplementary Table 3, and the counts are in the files deposited in Figshare). Further filtering within each family was done by LepMap3, to combine individuals with duplicate sequences that both passed the thresholds. No region(s) of the chromosome 12 assembly were found to be enriched with markers whose allele counts were low in our experiments, which might have revealed degenerated regions undetected by previous analyses

Different LepMap3 parameter values were necessary for different families, given their differences in numbers of sibships and total progeny numbers, and different levels of polymorphism (on which marker numbers depend). Parameters described in the LepMap3 documentation as being likely to be affected by such differences were therefore used in separate map estimates; when a change in one value improved the map, it was kept while another parameter value was changed to try and effect a further improvement. The results for a given family are similar with other parameter values. The parameter missingLimit=0.5 was used to filter out markers with different proportions of missing genotype data. For the Aripo river data, dataTolerance=0.0001 was used as the P-value threshold for excluding SNPs with non-Mendelian segregation ratios. For the Quare river data (with much larger numbers of markers, see below), a dataTolerance value of 0.04 was used, and an interference value of 0.03 was applied for the QLP1B1 data, and 0.02 for the QHPG5 family. Table S3 shows the informative markers called for each guppy chromosome assembly in each family.

Markers from all 23 guppy chromosomes were genotyped in the two Aripo river families. Linkage groups (LGs) with < 5 markers were removed using sizeLimit=5 for the

SeparateChromosomes module of LepMap3. Initial analysis with several values of the lodLimit parameter indicated that a value of 7 gave the number of linkage groups LGs closest to the guppy chromosome number for the LAH family data (18 chromosomes formed single LGs, while 5 were divided into 2 separate LGs; a total of 1,104 markers were mapped to these 28 LGs. These families, and two Quare river families, again from a low predation upstream and a high-predation downstream site, were also mapped with dense markers from the sex chromosome, in order to map the terminal PAR1 part of the chromosome in the sires. Chromosome 16 markers were also genotyped in the upstream Quare family (see the Results section).

Identifying the centromere ends of the guppy chromosomes

In male meiosis, many markers informative for mapping (across extensive regions of each telocentric chromosome) co-segregated in the families. As pericentromeric regions often have low recombination rates (Miller and Hawley 2017), large regions with low recombination in one or both sexes can identify the genetical centromere ends needed to assign the zero centimorgan locations in the genetic maps. To further verify the guppy chromosomes' centromere ends, we used GC content. This increases in regions with high recombination rates, due to GC-biased gene conversion (Marais and Galtier 2003). High recombination rates at guppy chromosome termini in male meiosis therefore lead to locally high GC, and previous analyses using the female reference assembly (based on short sequencing reads) detected this pattern. This shows that the centromeres are at the opposite ends (Charlesworth *et al.* 2020b). We repeated this to (i) use the improved assembly based on long-read sequencing of a male guppy (Fraser *et al.* 2020), and (ii) exclude possible enrichment with GC-rich transposable elements near the chromosome termini. We used RepeatModeler2 (Flynn *et al.* 2020) to generate an initial comprehensive set of candidate TE families, followed by RepeatMasker (Smit et al., http:// repeatmasker.org) analysis to make a library of consensus sequences from our set of TE families plus Teleostei data from Repbase (Bao *et al.* 2015). A filtered list of candidate repetitive sequences was manually annotated, following Goubert *et al.* (2022). Finally, RepeatMasker was run again to annotate repeat copies throughout the male reference genome. Files containing the consensus repeat library and locations of repeats in the male

reference genome were deposited in the Figshare repository (doi: 10.6084/m9.figshare.c.6644282). GC content in 20kb non-overlapping windows (Figure S2) was then estimated using a custom script that masked both the annotated repeats, and sites in the assembly other than A, T, G or C.

Finally, the candidate transposable element sequences were searched for types enriched near the ends of the chromosome assemblies. This revealed a sequence with multiple repeats concentrated at the ends identified as likely to be the centromere regions of each guppy chromosome, and much rarer in other genome regions (Table S2). Analysis using dotmatcher (https://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher, with the default settings, window size = 10, and threshold for identity = 23), suggested a repeat size of ~160-180 bp. Although guppy chromosomes are often described as acrocentric, our analysis the centromeres may not be at the very ends of some, or even all, the chromosomes (in other words, there may be one very short arm, see Figure S2).

Results

Crossover localization in guppy chromosomes: genetic maps of the autosomes

Female maps from two Aripo river families (Figure S4) confirm the previous conclusion using sparse microsatellite markers (Bergero *et al.* 2019) that crossovers occur evenly across the whole of each of the 23 guppy chromosomes, though several chromosomes could not be reliably mapped in the sibships from the low-predation site, due to insufficient SNPs. The ends of some chromosomes were not mapped in female meiosis (Figure S4). Some chromosome maps are therefore smaller than 50 cM, and the locations in female meiosis may be lower than the correct values. We therefore did not compare the map lengths in the two sexes, though our female and male maps generally suggest one crossover per bivalent (total map lengths around 50 cM), consistent with cytological estimates in males (Lisachov *et al.* 2015). However, the recombination pattern is very different between the sexes. In both families, our maps detect an extensive rarely recombining region at one end of each chromosome (usually at least 80% of the assembly, see Figure S4). Although previous results from chromosome 16 suggested that it might be an exception, our new mapping suggests crossover localization like that of the of

other guppy chromosomes (Figure S5). A previous high-density genetic map (Whiting *et al.* 2021) did not separate male and female meiosis, and the roughly constant recombination rate across chromosomes in female meiosis (see Figure 2A) obscures the relationships between physical positions and male meiotic map locations. However, high recombination rates are detectable near the chromosome termini, supporting terminal crossover localization in male meiosis.

The large regions in which crossovers are infrequent in the male maps are at the start of the assembly for 17 of the 23 guppy chromosomes, whereas GC content in introns and third codon positions is high at the opposite ends, suggesting that this reflects very high recombination rates in males, due to GC-biased gene conversion (Charlesworth *et al.* 2020b). To exclude the possibility that this pattern might instead reflect preferential presence of GC-rich transposable element(s) near the chromosome termini (Schield *et al.* 2022), we repeated the GC content analyses after masking repetitive sequences, and also identified a likely centromeric satellite sequence (see the Methods section). These analyses confirmed the centromeric ends previously identified (Figures S2 and S3) and identified their positions near one end of each chromosome (Table S2).

When SNPs near the centromere ends could be mapped in female meiosis, multiple autosomes appear to have large regions with few recombinants, in both Aripo river families (including chromosomes 6, 7 10, 11, 13, 14, 16, 20 and 21). The guppy may therefore have pericentromeric regions with low crossover rates like those in many other eukaryotes, in which centromere effects on recombination are common, as reviewed in Miller and Hawley (2017). Surprisingly, the maps in Figure S2 show no sign of any low recombination region on chromosome 2, which arose by a centric fusion between the homologues of the platyfish chromosomes Xm2 and Xm24 (Künstner *et al.* 2017), although its genetic map is longer than other chromosomes.

Crossovers on the guppy sex chromosomes

Figure 2A shows female and male maps of the guppy sex chromosome (chromosome 12) in sires from two Aripo and two Quare river families, again using high-throughput genotyping, but with targets in 500 chromosome 12 genes, to map as much of its assembly length as possible. Map locations close to zero should therefore be close to the correct values (unlike the situation for the autosomes). The female meiotic maps in the high- and low-predation families are clearly very similar (though the family from the Quare low-predation population includes only 23 progeny, and the female map is poorly estimated).

In the female assembly, a region between 10 and 20 Mb is inverted, probably reflecting an assembly error (corrected in the male assembly, see Fraser *et al.* 2020). The new maps support our previous conclusion for female meiosis, that the guppy X chromosome arrangement is always the same as that in the male assembly. The regions near both ends of the inverted region are repetitive, possibly explaining the assembly error. Of 17 genes in the region, one was not in the assembly of the platyfish homologue, Xm8, nor in any chromosome in the male assembly (Table S3C and D). SNPs were available in 9 of these genes (3 assigned to chromosome 17 in the male guppy assembly, one to chromosome 15, and 5 to chromosome 7; three are also not present in the Xm8 assembly). Markers in all 9 genes segregated autosomally in at least one guppy family (Table S3B, C and D), confirming that they are not carried on chromosome 12. The chromosome 7 genes include the *Olr1496*-like gene with a function in reproduction that shows higher coverage in guppy males than females, and a high F_{ST} value between the sexes in a guppy natural population sample (Lin *et al.* 2022).

In male meiosis, almost all crossovers are distal to 24.53 Mb in the male assembly, confirming presence of a small, but highly recombining, PAR1, while more centromere-proximal markers are completely sex linked in all the families studied (Figure 2B), with rare exceptions (described in the next section). Therefore, as indicated in Figure 1A, the centromere-proximal XY region is not completely sex linked, but includes a large PAR2. This recombines much more rarely in males than females, and therefore does not simply reflect pericentromeric rarity of recombination. The XY region must include this PAR2 plus any completely Y-linked region (which could potentially be an extensive MSY region, as shown in Figure 1A). Detailed segregation results for all four families are shown in Supplementary Tables S3A to D, and Table

S4 summarizes results from all families with results for microsatellite and/or SNP markers distal to 20 Mb in either the male or female assemblies.

Y-X recombinants

In addition to two recombinant individuals described previously (Charlesworth *et al.* 2020a), a further family with a recombinant male, indicating a crossover in the XY (or PAR2) region, was found among the total of 669 progeny genotyped. The recombinant male fish had a gonopodium and looked like other males, unlike the recombinant female, which resembled other females in her family (and other families). Our combined results suggest that the male-determining factor is within a small physical region in the female assembly. Figure 3 and Table 1 summarize the results from the three recombinant individuals, two of them with many more SNP markers than previously mapped. Surprisingly, all three crossover events occurred near 21 Mb in the female assembly (which provided the finest detail, as positions the male assembly are known only in terms of the genes in which mapped markers are located, not the base position in the assembly). This region is in a similar position in the male assembly, though the male and female assemblies differ somewhat (see below). The recombinant male in the QHPG5 family appears to have two further crossovers in more distal positions. A double recombinant is unlikely, given the rarity of crossovers in males, but an inversion in the region in this sire could produce the observed pattern (Wall *et al.* 2022). This possible rearrangement does not affect the conclusions below about candidate male-determining genes.

The new recombinant individual (M7 in sibship 2 of the QLP1B1 family in Table S3D), inherited his sire's paternal X-linked alleles (found in all 14 female and none of the 6 male full sibs) for the centromere-proximal 1,660 SNPs informative in male meiosis (omitting SNPs that map to chromosome 7 in the male assembly, and 36 SNPs with evidence of duplications or other unexpected segregation patterns); M7's genotypes then change to the sire's Y-linked alleles until the PAR1 boundary (the change occurs near 20,930,721 bp, see Table 1, based on Supplementary Table 3D). The male-determining factor must therefore be distal to the informative marker at 20,698,741 bp. The recombinant female previously found in the PMLPB2

family (Supplementary Table 3E) has an almost exactly opposite pattern, inheriting its sire's proximal X-linked alleles until close to 20 Mb (after correcting the assembly error in the female assembly), and Y-linked alleles for three more distal markers that show full sex-linkage in all other progeny. As the recombinant individual is female, the male-determining factor must be proximal to the first such marker, which is at 21,363,193. This locates the M factor within the 664,452 bp interval between this position and the most proximal paternal X-linked allele in male M7 in the QLP1B1 family.

The interval identified includes only 24 genes in the female assembly annotation (Table S5). Twelve of these 24 sequences are not in the male chromosome 12 assembly, and those that are present are assembled in three different chromosome 12 regions (indicated by three different colours in column G of Table S5). Of the markers mapped in QLP1B1 dams, 6 in the *ptpn13* gene map near 25 cM, while 19 markers (in 9 other genes) map near 50 cM (Figure S6); although only two of these are assembled on chr12, they co-segregate with the male-determining locus in male meiosis, and are therefore not autosomal.

Four genes belonging to the interval identified are near 5 Mb in the male assembly, close to Contig IV, a sex-determining candidate region which includes repetitive sequences that are duplicated in the more distal Contig XII, near 25 Mb in the male assembly (Fraser *et al.* 2020). We mapped markers in three Contig IV genes (though several show signs of being duplicated, see Figure S7B and Table S3), and a microsatellite marker, chr12cIV_AC618 (the only one that yielded reliable genotypes, out of 11 tested in this region, see Figure S7A and B). This microsatellite co-segregated with sex in the sires of three families (LAH, ALP2B2 and QHPG5), as did the few reliable genic markers. In female meiosis, the microsatellite maps (in LAH and ALP2B2 dams) near much more distal chr12 markers (not near SNPs in genes assembled near 5 Mb), and the Contig IV gene cyclin 1 maps (in the QHPG1 dam) near 50 cM. These results, albeit confusing, are nevertheless consistent with a male-determining factor located in a small region near Contig XII (see Discussion).

The sex chromosome terminal region and PAR1

Figure 2B shows genetic map results for just the terminal part of the sex chromosome, from the same high-throughput SNP genotyping experiments and families as in Figure 2A. Markers in several scaffolds that are unplaced in the female or male assembly co-segregate with sex-linked markers (Fraser *et al.* 2020), and can be assigned terminal positions based on the LAH and/or ALP2B2 families (Supplementary Tables 3A and B and Figure S8). Scaffolds NW_007615023.1 and NW_00761531.1 are unplaced in the female assembly, but include genes with homologous sequences near the end of the platyfish Xm8 assembly most distant from the centromeres, and map genetically within PAR1 (Table S4). However, markers in unplaced contigs were difficult to map, owing to duplications and/or low coverage or repetitive sequences (reminiscent of Morgan et al.'s 2019 findings in mouse sub-telomeric regions). One marker, 023_AC17, was duplicated in some individuals, but only in the ALP2B2 family.

Even using the male assembly, which is based on long-read sequences, including PacBio and Hi-C analyses (Fraser *et al.* 2020), the PAR1 genetic map locations are in the opposite order from the physical positions of the markers (Supplementary Tables S4 and S5). Moreover, several markers from the end of the male assembly did not recombine with the male-determining locus in any of the four families (Figure 2B and Supplementary Tables S3A to D). As assembly problems and duplications are possible, the most informative markers for comparing the PAR1 boundary positions in different families are the most centromere-proximal ones that showed complete Y-linkage in the Experiment 3 and 4 sires (with the highest density of chromosome 12 markers). Wherever information is available, such markers are found just to the left of 24.5 Mb in the male assembly, and all sires yielded very similar maps in this region (Table S6 summarizes results from all the families).

Recombination in a probable XX male

One family from a Guanapo low-predation site (GLPGrp1 in Table S1) produced a large number of progeny with a highly female-biased sex ratio. Genotypes of the parents and progeny for four informative microsatellite markers indicated that 8 male and 7 female progeny derived from sire 1, while alleles at all four marker loci that were absent from sire 1 uniquely identified

another sire, 2, whose 63 progeny were all female. This was probably an XX male. A sex change is not implausible, as such changes have been observed in guppies (Winge 1930). Sire 2's alleles segregated in 1:1 ratios in the progeny of two different dams (36 from one dam, and 27 from the other), suggesting that the biased sex rario was not due to meiotic drive in an XY male.

This individual allowed us to ask whether the crossover pattern depends on the sex phenotype in guppies. Rather than crossovers occurring across the whole chromosome (as described above for other XX females), only one crossover event was detected in sire 2's progeny, between markers *rgrfAC* and *AG179* (respectively at 3.78 and 17.36 Mb in the male assembly). In the ALP2B2 and LAH family dams, the map locations of the *rgrfAC* and *AG179* markers are around 35 and 21.5 cM apart, respectively (Table S3A and B). We might thus expect at least 10 recombinants among the 63 genotyped progeny. The observed much lower number (Table S3F) suggests that this phenotypic male's crossover pattern is much more similar to that of XY males, with no, or rare, crossovers proximal to 24 Mb. Although one crossover was observed in this family, this could reflect the occasional events that occur in male meiosis within the XY region (see above).

Discussion

Crossover patterns in guppy male and female meiosis

Our genetic mapping results support the previous evidence that, specifically in male meiosis, crossovers are highly localized at the non-centromere ends of all 23 guppy chromosomes (Bergero *et al.* 2019; Charlesworth *et al.* 2020a). Similar findings have been reported in some other fish where genome sequencing has allowed genetic maps to be compared with physical maps of the chromosomes (reviewed in Sardell and Kirkpatrick 2020). The pattern is pronounced in Atlantic halibut (Edvardsen *et al.* 2022), though other species, such as Atlantic herring (Pettersson *et al.* 2019), show much less crossover localization and smaller sexual dimorphism in crossover patterns and rates. Even when crossovers are localized, as in the guppy, they are not completely restricted to terminal regions of the chromosomes, as our results support previous inferences of rare recombination proximal to PAR1 in XY males.

Our main question is whether recombination suppression has evolved recently between the guppy XY pair. One possible test is to ask whether the crossover localization in male guppies is more extreme for the sex chromosome pair than for the autosomes. In male meiosis, our two Aripo river families often reliably indicated crossovers occurring as far as 20% of the physical distance from the distal end of an autosome (though terminal markers on chromosomes 11 and 18 did not reveal crossovers in the sires of either family; Figure S4), whereas almost all crossovers on the XY pair were in the terminal 5% (Figure 2). However, the autosomal results often differ between the two families. For example, chromosome 5 appears to be well mapped in both families, and multiple terminal markers indicate crossovers in the high-predation family (the left-hand columns in these figures), but not in the low-predation family. Since many more progeny were genotyped in the latter, this is probably a real difference between the families, which could reflect genetic differences or effects of the conditions under which the sires were raised. Recombination rate estimates from one guppy sire (or sires from one population, or experiment) may not apply to all other guppies; such differences have been documented in other species, including a fish (Kivikoski *et al.* 2023). Overall, therefore, we cannot currently conclude whether crossover localization is stronger for the sex chromosome pair than the autosomes.

It is, however, clear that the sex chromosome recombination patterns do not greatly differ between the uptream and downstream males from either river studied (Figure 2). First, the lack of recombinants across much of chromosome 12 in the progeny from the low-predation site sires is not due to small family sizes, but reflects a large region that recombines rarely (an XY region, see Figure 1A). Figure 2B shows that this occupies very similar parts of chromosome 12 in our upstream and downstream sires from both rivers. Second, the only three crossover events detected are distal to 20 Mb. Therefore the XY region probably does not include an extensive MSY that includes old and young strata (as well as a more recombinationally active region centromere-proximal to the younger strata) as Wright *et al.* (2017) suggested.

We cannot, however, reliably determine the PAR1 size. The physical position where the male genetic map positions change sharply from near zero to around 50 cM clearly cannot be

the MSY-PAR1 boundary, as genetic map positions decline towards the chromosome terminus, the opposite of what is expected. The entire distal chromosome 12 region thus appears to be inverted, even in the male assembly (Figure 2B). Several terminal markers co-segregate with the sex-determining locus in our families. These could be within PAR1 sequences physically near the boundary with the XY region. Alternatively, they could belong to an XY region that has been mis-assembled at the chromosome terminus (see Figures 1A and 3), in which case PAR1 is smaller than the region distal to the sharp change in genetic map locations. The Y and X sizes in the current male assembly are very similar (Fraser *et al.* 2020). This assembly therefore does not suggest the presence of any extensive fully Y-linked region. Although an MSY, or old stratum like that suggested by Wright *et al.* (2017), cannot be excluded, this result, and those summarized in Figure 3, suggest that any such region is physically smaller than 3 Mb.

Within PAR1, the very high crossover rates make reliable genetic mapping possible. The relationship between the genetic and physical map positions appears to be roughly linear, suggesting uniform crossover probabilities across the region, as in Figure 1C. However, if, as just discussed, the markers assembled most terminally belong in a rarely recombining XY region, and not in PAR1, a hotspot could be present near the boundary between the two regions, as has been found in the ostrich (Yazdi *et al.* 2023). Figure S9 evaluates this possibility further, and shows a gap in the genetic map locations of the terminal markers and those for the markers that are certainly within PAR1. The data are also consistent with an assembly error, which can be resolved when the physical arrangement can be reliably determined. However, this may be difficult if the terminal region includes repetitive sequences and/or rearrangements, as in mouse sub-telomeric regions (Morgan *et al.* 2019). In the mouse, the GC content at neutral sites in sequences changes sharply enough to identify the PAR1 boundary (Marais and Galtier 2003). In the guppy, however, although the change is moderately sharp, it does not identify the boundary precisely (Charlesworth *et al.* 2020b).

Associations between coloration traits and the Y-linked region

The very similar recombination patterns in males from upstream and downstream sites appears to conflict with the evidence outlined in the Introduction, suggesting that male coloration traits may show closer linkage to the male-determining locus in downstream than upstream populations (where predation pressure is less severe, and coloration is less disadvantageous). However, there is direct evidence for recombination differences only for the *Sb* factor (which shows near-complete Y-linkage). The indirect experiments using testosterone-treated females (Haskins *et al.* 1961; Gordon *et al.* 2012) yielded results consistent with coloration being inherited in a non-Y-linked manner most often in females from upstream sites (see Introduction). This does not, however, imply that Y chromosome crossover rates for partially sex-linked genes differ between these populations. There is thus no contradiction with our results, which assessed crossing over for markers at different distances from the PAR1 boundary. However, we identified few SNPs with crossover rates less than 20% from the boundary (Figure 2B), and, with our family sizes, we could not reliably have detected differences between very small recombination rates, like those between the *Sb* coloration factor and the male-determining locus.

Moreover, other explanations for the observations of Gordon et al. (2012) are not excluded. As explained in the Introduction, SA factors unlinked to the male-determining locus (either autosomal or X-linked, like the non-Y-linked factors inferred by Morris *et al.* 2020) should often become fixed within populations, rather than establishing polymorphisms. All or most testosterone-treated females would then display the trait, depending on the penetrance. Fixation is especially likely in upstream environments where predation pressure selecting against females showing such traits is weak, and recent bottlenecks (Qiu *et al.* 2022) could have allowed loss of low frequency alleles by genetic drift. The previous observations can therefore be reconciled with the evidence that coloration traits are composite phenotypes affected by multiple factors, not all Y-linked. The maintenance of the *Sb* factor polymorphism in upstream sites probably reflects this trait's simple Y-linked inheritance and likely advantage when rare, which can maintain Y-linked polymorphism (Clark 1987). Overall, therefore, the fact that few females from downstream sites carry coloration factors could simply reflect fewer high

frequency unlinked coloration factors, not closer linkage between individual partially Y-linked factors and the male-determining locus (the information that is provided by mapping SNPs).

A possible location for the male-determining factor

Based on the female assembly, the observed X-Y crossovers suggest that the male-determining factor is near 21 Mb (Figure 3, Table 1). This is not close to PAR1, which is more distal (see above). In the family with the highest marker density, QLP1B1, markers close to the crossover points are in a repetitive region near one end of the erroneous inversion in the female assembly (Fraser *et al.* 2020). The female genetic map positions of these markers fall into two sets, consistent with some markers being within the inversion near its distal end, and others being just distal to it (Figure S6). In the male assembly, some of them are at the terminal end of chromosome 12, consistent with our female genetic map results; as discussed above, markers in this region of the male assembly generally co-segregate with sex, but their true positions could be between 24 and 25 Mb, just distal to a region in which the male and female assemblies are colinear (Figure 3, Figure S9). Others are assembled near 11 Mb, and again some of their female genetic map positions are consistent with this physical position, but others co-segregate with markers some distance proximal to PAR1.

These results are consistent with the previous suggestion that Contig IV sequences, or duplicates of them in this region, might include the male-determining factor (Fraser *et al.* 2020), and with associations of sex-linked molecular variants with individuals' sex phenotypes in natural guppy populations. Associations (albeit incomplete ones) are consistently detected for SNPs in two chromosome 12 regions proximal to the boundary between PAR1 and the rest of the chromosome (Charlesworth *et al.* 2020a; Fraser *et al.* 2020; Almeida *et al.* 2021; Qiu *et al.* 2022). One region coincides with the region near 21 Mb where our recombinants suggest that the male-determining factor may be located.

The second region of LD with the male-determining factor is near 25 Mb, and may reflect an assembly error. Alternatively, it could indicate the presence of two occasionally recombining Y-linked loci with balanced polymorphisms. Under extremely close linkage with

the maleness factor, an SA polymorphism can lead to high Y-X differences for molecular variants, with a lesser signal between the two loci (Kirkpatrick and Guerrero 2014). Recombination in this XY region (between 21 Mb and the PAR1 boundary, see Figure 3) may be infrequent enough to maintain two peaks of molecular polymorphism, one near a male-determining factor near 21 Mb, and another near a SA polymorphism locus, such as a male coloration factor.

Coloration factors are not the only Y-X differences expected to cluster near the completely Y-linked region. SA mutations conferring other male-specific benefits should be enriched within a small linkage distance from the male-determining locus, by the sieve effect outlined in the Introduction. Without reliably segregating major effect male coloration or other SA factors, it is difficult to test this. It is currently unclear how many Y-linked polymorphisms are segregating in natural guppy populations, even for just male coloration traits. If there are multiple SA polymorphisms, LD caused by close linkage with the male-determining factor could produce a complex situation that might make it difficult to pinpoint their positions, as different males would carry different factors, at different distances from the male-determining factor. This could explain the slightly different regions in which molecular markers show associations with the sexes in different populations in the studies just mentioned.

Coloration genes and other SA polymorphisms are expected to be closely linked to the M factor, either within an MSY region or physically near the PAR1 boundary (see the Introduction). This is consistent with direct genetic evidence that all partially Y-linked coloration factors appear to be within 10 cM of the M factor (Winge and Ditlevsen 1947). Although such polymorphisms would create selection for closer linkage with the sex-determining locus, the close linkage that is required for SA polymorphisms to establish (and fixation to be avoided), implies that changes in recombination could be undetectably small.

Sexual dimorphism in gene expression

Overall, our result that the recombination pattern on the guppy Y chromosome does not differ between high and low predation populations is not compatible with the hypothesis that

recombination is currently becoming suppressed. This does not imply that SA mutations were unimportant in guppy evolution. Male-specific expression of coloration traits is an equally interesting consequence of SA selection. Unlike the evolution of recombination rates, polymorphism is not required. SA mutations that become fixed or reach high frequencies (including variants that establish polymorphisms due to diminishing advantages as they become frequent in populations) may evolve male-limited expression, to remove the disadvantages in females. The guppy Y does not appear to carry many sequences that are missing from the X, and certainly not protein-coding genes, based on the current assemblies (Fraser *et al.* 2020). Male coloration factors are therefore probably not Y-specific sequences, and at least some can be carried on the X after crossing over (Winge and Ditlevsen 1947).

In guppy males, the coloration traits develop at maturity, so their expression probably requires testosterone. The coloration mutations' effects could therefore have been completely or largely male-limited when they first arose, in which case there is no need to consider how male-specific expression evolved. Evolution of male-specificity is also unlikely for mutations that arose in genes with complete or strong sex linkage, because selection against expression in females would be weak if crossovers very rarely produce female carriers. However, complete male-limitation of expression of the initial mutations cannot explain the apparent under-representation of autosomal factors. The common observation of Y-linkage of the polymorphic factors therefore supports the action of a selective sieve, with autosomal mutations expressed in females often being eliminated or fixed. Male-specific expression must subsequently have evolved for autosomal factors that are no longer expressed in females, and similar changes may also have affected expression of traits controlled by sex-linked factors. Selection in males might lead to stronger or more reliable expression in the presence of testosterone, via evolution of cis-acting factors perhaps increasing dominance levels of coloration factors. Trans-acting autosomal factors might also have suppressed expression of coloration traits in occasional females that inherited the mutations, for example ensuring expression of such factors only in the presence of testosterone, while still allowing expression in heterozygous males. In the future, it will be interesting to test whether such changes are detectable.

Data availability. The genotype data and the files from the analysis of repetitive sequences have been deposited in Figshare (doi: 10.6084/m9.figshare.c.6644282).

Acknowledgements: We thank colleagues A. Wilson and D.P. Croft at Exeter University for field collection of fish used to make our families and maintenance of the families, and I.W. Ramnarine and R. Mahabir (University of the West Indies, Trinidad and Tobago) for assistance with the field collection.

Funding: This project was supported by European Research Council (ERC) grant number 695225 (GUPPYSEX). M. Konczal was supported by the Polish National Agency for Academic Exchange (Bekker Programme) and Polish National Science Centre (grant number: UMO-2018/31/D/NZ8/00091).

References

Almeida, P., B. A. Sandkam, J. Morris, I. Darolti, F. Breden *et al.*, 2021 Divergence and remarkable diversity of the Y chromosome in guppies. Molecular Biology and Evolution 38**:** 619–633. doi: 10.1093/molbev/msaa257

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. doi:

Avvaru, A. K., D. Sowpati and R. Mishra, 2018 PERF: an exhaustive algorithm for ultra-fast and efficient identification of microsatellites from large DNA sequences Bioinformatics 34**:** 943–948. doi: 10.1093/bioinformatics/btx721

Bao, W., K. K. Kojima and O. Kohany, 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6**:** 11. doi: 10.1186/s13100-015-0041-9

Bergero, R., J. Gardner, B. Bader, L. Yong and D. Charlesworth, 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. doi: 10.1073/pnas.1818486116

Charlesworth, B., and D. Charlesworth, 1978 A model for the evolution of dioecy and gynodioecy. American Naturalist 112**:** 975-997. doi:

Charlesworth, D., R. Bergero, C. Graham, J. Gardner and K. Keegan, 2021 How did the guppy Y chromosome evolve? PLOS GENETICS 17**:** e1009704. doi: 10.1371/journal.pgen.1009704

Charlesworth, D., R. Bergero, C. Graham, J. Gardner and L. Yong, 2020a Locating the sex determining region of linkage group 12 of guppy (*Poecilia reticulata*). G3 (Bethesda) 10**:** 3639-3649. doi: 10.1534/g3.120.401573

Charlesworth, D., and B. Charlesworth, 1975 Theoretical genetics of Batesian mimicry. II. Evolution of supergenes. J. Theoret. Biol. 55**:** 305-324. doi:

Charlesworth, D., and B. Charlesworth, 1980 Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genetical Research (Cambridge) 35**:** 205-214. doi: 10.1017/S0016672300014051

Charlesworth, D., Y. Zhang, R. Bergero, C. Graham, J. Gardner *et al.*, 2020b Using GC content to compare recombination patterns on the sex chromosomes and autosomes of the guppy, *Poecilia reticulata*, and its close outgroup species. Molecular Biology and Evolution 37**:** 3550–3562. doi: 10.1093/molbev/msaa187

Clark, A., 1987 Natural selection and Y-linked polymorphism. Genetics 115**:** 569-577. doi:

Edvardsen, R. B., O. Wallerman, T. Furmanek, L. Kleppe, P. Jern *et al.*, 2022 Heterochiasmy and the establishment of gsdf as a novel sex determining gene in Atlantic halibut. PLOS Genetics 18**:** e1010011. doi: 10.1371/journal.pgen.1010011

Fisher, R. A., 1931 The evolution of dominance. Biol. Rev. 6**:** 345-368. doi:

Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark *et al.*, 2020 RepeatModeler2: automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences of the United States of America 117**:** 9451-9457. doi: 10.1073/pnas.1921046117

Fraser, B. A., J. R. Whiting, J. R. Paris, C. J. Weadick, P. J. Parsons *et al.*, 2020 Improved reference genome uncovers novel sex-linked regions in the guppy (*Poecilia reticulata)*. Genome Biology and Evolution 12**:** 1789–1805. doi: 10.1093/gbe/evaa187

Garrison, E., and G. Marth, 2012 Haplotype-based variant detection from short-read sequencing. arXiv**:** arXiv:1207.3907. doi:

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

Goubert, C., R. J. Craig, A. F. Bilat, V. Peona, A. A. Vogan *et al.*, 2022 A beginner's guide to manual curation of transposable elements. Mob DNA 13**:** 7. doi: 10.1186/s13100-021-00259-7

Haskins, C., and E. F. Haskins, 1951 The inheritance of certain color patterns in wild populations of *Lebistes reticulatus* inTrinidad. Evolution 5**:** 216-225. doi: 10.2307/2405461

Haskins, C., E. F. Haskins, J. McLaughlin and R. E. Hewitt, 1961 Polymorphisms and population structure in *Lebistes reticulatus*, an ecological study, pp. 320-395 in *Vertebrate Speciation*, edited by W. F. Blair. University of Texas Press, Austin, TX.

Jordan, C., and D. Charlesworth, 2012 The potential for sexually antagonistic polymorphism in different genome regions. Evolution 66**:** 505–516. doi: 10.1111/j.1558-5646.2011.01448.x

Khoo, G., T. Lim, W. Chan and V. Phang, 1999 Linkage analysis and mapping of three sex-linked color pattern genes in the guppy, *Poecilia reticulata*. Zoological Science 16**:** 893-903. doi: 10.2108/zsj.16.893

Kirkpatrick, M., G. Dixon, J. Sardell, M. Schartl, M. P. Josephson *et al.*, 2022 Evolution of the canonical sex chromosomes of the guppy and relatives. G3-Genes Genomes Genetics 12**:** jkab435. doi: 10.1093/g3journal/jkab435

Kirkpatrick, M., and R. Guerrero, 2014 Signatures of sex-antagonistic selection on recombining sex chromosomes. Genetics 197**:** 531-541. doi: 10.1534/genetics.113.156026

Kivikoski, M., A. Fraimout, P. Rastas, A. Löytynoja and J. Merila, 2023 Repeatability of crossover rate in wild sticklebacks. Biological Journal of the Linnean Society 10.1093/biolinnean/blad032. doi: 10.1093/biolinnean/blad032

Künstner, A., M. Hoffmann, B. A. Fraser, V. A. Kottler, E. Sharma *et al.*, 2017 The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. PLOS ONE 11**:** e0169087. doi: 10.1371/journal.pone.0169087

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science 286**:** 964-967. doi: 10.1126/science.286.5441.964

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv arXiv:1303.3997. doi:

Lin, Y., I. Darolti, B. L. S. Furman, P. Almeida, B. A. Sandkam *et al.*, 2022 Gene duplication to the Y chromosome in Trinidadian Guppies. Molecular Ecology 31**:** 1853-1863. doi: https://doi.org/10.1111/mec.16355

Lindholm, A., and F. Breden, 2002 Sex chromosomes and sexual selection in Poeciliid fishes. American Naturalist 160**:** S214-S224. doi: 10.1086/342898

Lisachov, A., K. Zadesenets, N. Rubtsov and P. Borodin, 2015 Sex chromosome synapsis and recombination in male guppies. Zebrafish 12**:** 174-180. doi: 10.1089/zeb.2014.1000

Marais, G., and N. Galtier, 2003 Sex chromosomes: how X-Y recombination stops. Curr. Biol. 13**:** R641-643. doi: 10.1016/S0960-9822(03)00570-0

Miller, D., and R. Hawley, 2017 Meiotic recombination: taking the path less traveled. Current Biology 27**:** R19–R41. doi: 10.1016/j.cub.2016.11.002

Morgan, A. P., T. A. Bell, J. J. Crowley and F. P.-M. d. Villena, 2019 Instability of the pseudoautosomal boundary in house mice. Genetics 212**:** 469-487. doi: 10.1534/genetics.119.302232

Morris, J., I. Darolti, W. v. d. Bijl and J. Mank, 2020 High-resolution characterization of male ornamentation and re-evaluation of sex linkage in guppies. Proceedings of the Royal Society B 287**:** 20201677. doi: 10.1098/rspb.2020.1677

Pettersson, M. E., C. M. Rochus, F. Han, J. Chen, J. Hill *et al.*, 2019 A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. Genome Res 29**:** 1919-1928. doi: 10.1101/gr.253435.119

Postma, E., N. Spyrou, L. Rollins and R. C. Brooks, 2011 Sex-dependent selection differentially shapes genetic variation on and off the guppy Y chromosome. Evolution 65**:** 2145–2156. doi: 10.1111/j.1558-5646.2011.01314.x

Potter, T., J. Arendt, R. D. Bassar, B. Watson, P. Bentzen *et al.*, 2023 Female preference for rare males is maintained by indirect selection in Trinidadian guppies. Science 380**:** 309-312. doi: 10.1126/science.ade5671

Qiu, S., L. Yong, A. Wilson, D. P. Croft, C. Graham *et al.*, 2022 Partial sex linkage and linkage disequilibrium on the guppy sex chromosome. Molecular Ecology 31**:** 5524– 5537. doi: 10.1111/mec.16674

Rastas, P., 2017 Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33**:** 3726-3732. doi: 10.1093/bioinformatics/btx494

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. doi:

Rouyer, F., M. Simmler, C. Johnsson, G. Vergnaud, H. Cooke *et al.*, 1986 A gradient of sex linkage in the pseudoautosomal region of the human sex-chromosomes. Nature 319**:** 291-295. doi: 10.1038/319291a0

Sardell, J., and M. Kirkpatrick, 2020 Sex differences in the recombination landscape. American Naturalist 195: 361–379. doi: 10.1086/704943

Schield, D. R., B. W. Perry, D. C. Card, G. I. M. Pasquesi, A. K. Westfall *et al.*, 2022 The rattlesnake W chromosome: A GC-rich retroelement refugium with retained gene function across ancient evolutionary strata. Genome Biology and Evolution 14. doi: 10.1093/gbe/evac116

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

Smeds, L., T. Kawakami, R. Burri, B. P, A. Husby *et al.*, 2014 Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. Nature Communications 5: 5448. doi: 10.1038/ncomms6448

Tripathi, N., M. Hoffmann, E. Willing, C. Lanz, D. Weigel *et al.*, 2009 Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. Proceedings of the Royal Society B 276: 2195-2208. doi: 10.1098/rspb.2008.1930

Turner, J., 1977 Butterfly mimicry: the genetical evolution of an adaptation. Evol. Biol. 10: 163-206. doi:

Wall, J. D., J. A. Robinson and L. A. Cox, 2022 High-resolution estimates of crossover and noncrossover recombination from a captive baboon colony. Genome Biology and Evolution 14. doi: 10.1093/gbe/evac040

Whiting, J. R., J. R. Paris, M. J. v. d. Zee, P. J. Parsons, D. Weigel *et al.*, 2021 Drainage-structuring of ancestral variation and a common functional pathway shape limited genomic convergence in natural high- and low-predation guppies. PLoS Genetics 10.1371/journal.pgen.1009566: e1009566. doi: 10.1371/journal.pgen.1009566

Winge, O., 1927 The location of eighteen genes in *Lebistes reticulatus*. Journal of Genetics 18: 201-217. doi:

Winge, O., 1930 On the occurrence of XX males in *Lebistes* with some remarks on Aida's so called 'non-disjunctional males' in *Aplocheilus*. Journal of Genetics 23: 69–76. doi:

Winge, O., and E. Ditlevsen, 1947 Colour inheritance and sex determination in *Lebistes*. Heredity 1: 65-83. doi: 10.1038/hdy.1947.4

Wright, A., I. Darolti, N. Bloch, V. Oostra, B. Sandkam *et al.*, 2017 Convergent recombination suppression suggests a role of sexual conflict in guppy sex chromosome formation Nature Communications 8: 14251. doi: 10.1038/ncomms14251

Yazdi, H. P., C. Olito, T. Kawakami, P. Unneberg, M. F. Schou *et al.*, 2023 The evolutionary maintenance of ancient recombining sex chromosomes in the ostrich. PLOS Genetics 19: e1010801. doi: 10.1371/journal.pgen.1010801

Figure legends

Figure 1. Two situations that could increase the terminal localization of crossovers, creating closer linkage between the M factor (vertical bar labelled M) and coloration factors within the terminal chromosome region (whose locations are discussed in the text). **A**. Organisation of the

guppy Y chromosome, with a male-determining factor within an "XY region" and a terminal "PAR1" with a high recombination rate. Crossovers may occasionally occur in the XY region (making it a partially sex-linked "PAR2", see text), or it may be an extended completely sex-linked region, or "MSY" (thick horizontal box). B and C diagram possible recombination changes, with dashed lines showing ancestral states, and unbroken lines changed states. **B**. An ancestral hotspot within PAR1, near the boundary with the XY region moves to a more distal position. **C**. The PAR1 boundary moves, shrinking PAR1 and expanding the XY region.

Figure 2. Genetic map results from parents of four families for chromosome 12. **A**. Male maps (square symbols) and female maps (round symbols) for the whole of the sex chromosome pair in each family indicated. The centromere position is at the left-hand end. **B**. Maps from the male parents of the same families, as indicated in the key, for the terminal part of the guppy sex chromosome, starting from 21 Mb in the male assembly, showing the very similar maps in the males from parents from high-and low-predation parents (HP and LP, respectively, in the key). B also shows that markers proximal to about 24.5 Mb are not part of the highly recombining pseudo-autosomal region, PAR1, in these families.

Figure 3. Diagram summarizing the recombinant progeny individuals found in three families and the position of the maleness factor inferred from them within the XY region. The rows at the top indicate the regions shown in Figure 1A, and the regions with major assembly errors and high repeat content mentioned in the text. The colinear region is shown in Figure S9. Each cell in the body of the table indicates whether the recombinant individual inherited its sire's X- or Y-linked allele, which could be inferred from the sexes that inherited each allele in the recombinant individual's siblings, together with the parent genotypes. The positions of the genotyped markers are shown in Supplementary Tables 3 C and D for the two Quare river families with large numbers of SNPs genotyped, and in Supplementary Table 3E for the much smaller numbers of markers genotyped in the Petit Marianne family. As discussed in the text, the apparent inconsistency in the QHPG5 family could reflect an inversion.

Table 1. SNP positions that are informative about the recombination events in the QHPG5, QLPB1 and PMLPB2 families. The results suggest that the male-determiner is within Region 2a. The positions of the markers are in Mb in the female assembly, where known, because this assembly is annotated, but the positions are mostly very similar in the male assembly, except for the error in the female assembly, where sequences between about 10 and 20 Mb are inverted (Supplementary Table 3E includes positions in both assemblies for the PMLPB2 family, where this is important).

QHPG5 sibship 2 recombinant male				QLP1B1 sibship 2 recombinant male				PMLPB2 sibship 1 recombinant female			
Region (paternal alleles)	**First site**	**Last site**	**Size of region (Mb)**	**Region (paternal alleles)**	**First site**	**Last site**	**Size of region (Mb)**	**Region (paternal alleles)**	**First site**	**Last site**	**Size of region (Mb)**
1 (X)	0.13	20.60	20.47	1 (X)	0.13	20.70	20.83	1 (X)	1.27	11,748,712 [1]	~ 20
2a (Y)	21.05	22.90	1.85								
2b (X)	24.12	24.51	0.39	2 (Y)	20.93	25.38	4.45	2 (Y)	22.76	25.82	3.06
2c (Y)	24.85	25.38	0.53								
PAR1	25.52	26.40	0.88	PAR1	25.52	26.43	0.91	No PAR1 markers mapped			

[1] This is close to the end of the error in the female assembly, and the position is near 20 Mb in the male assembly.

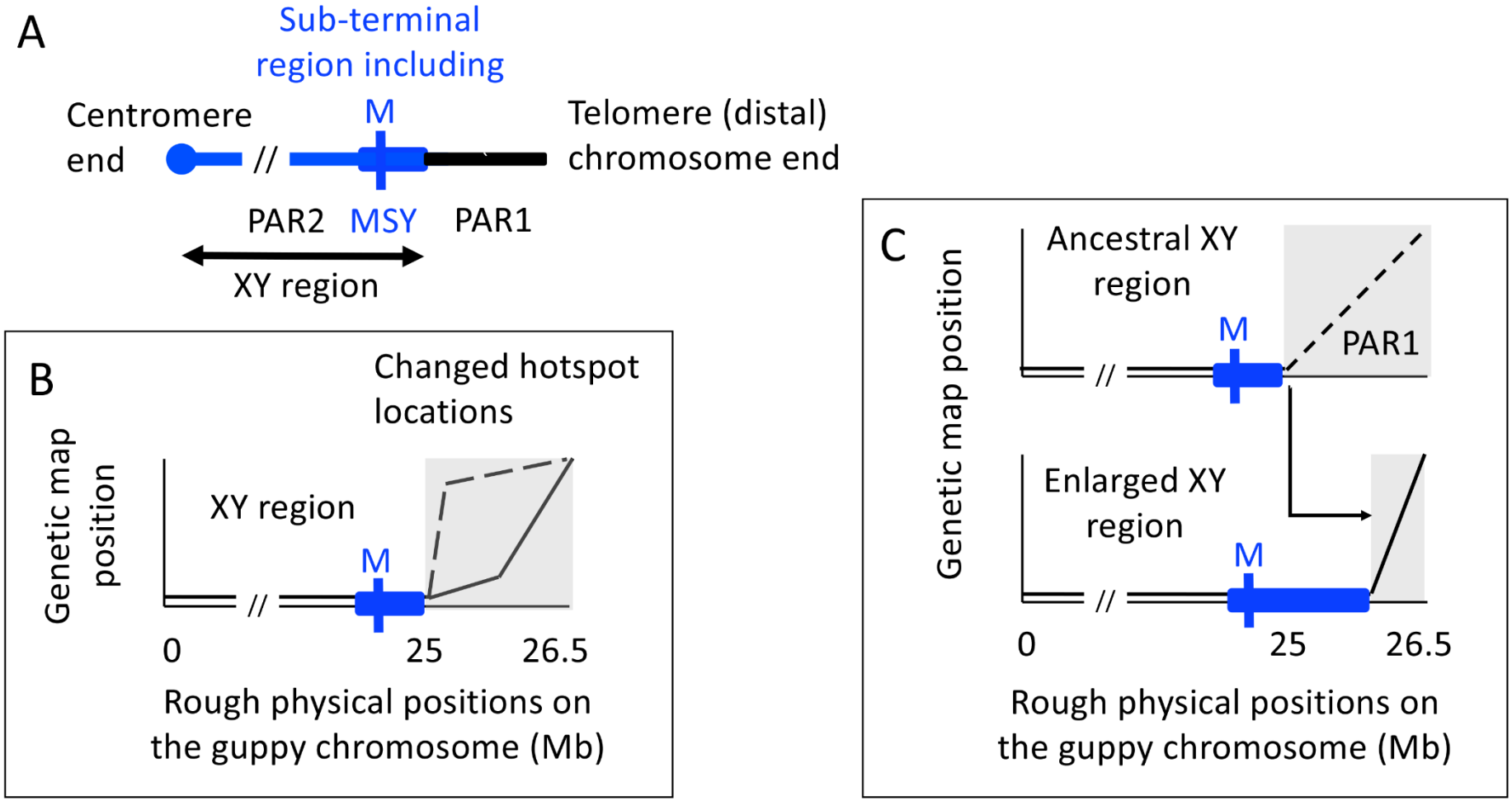

Figure 1
246x139 mm (x DPI)

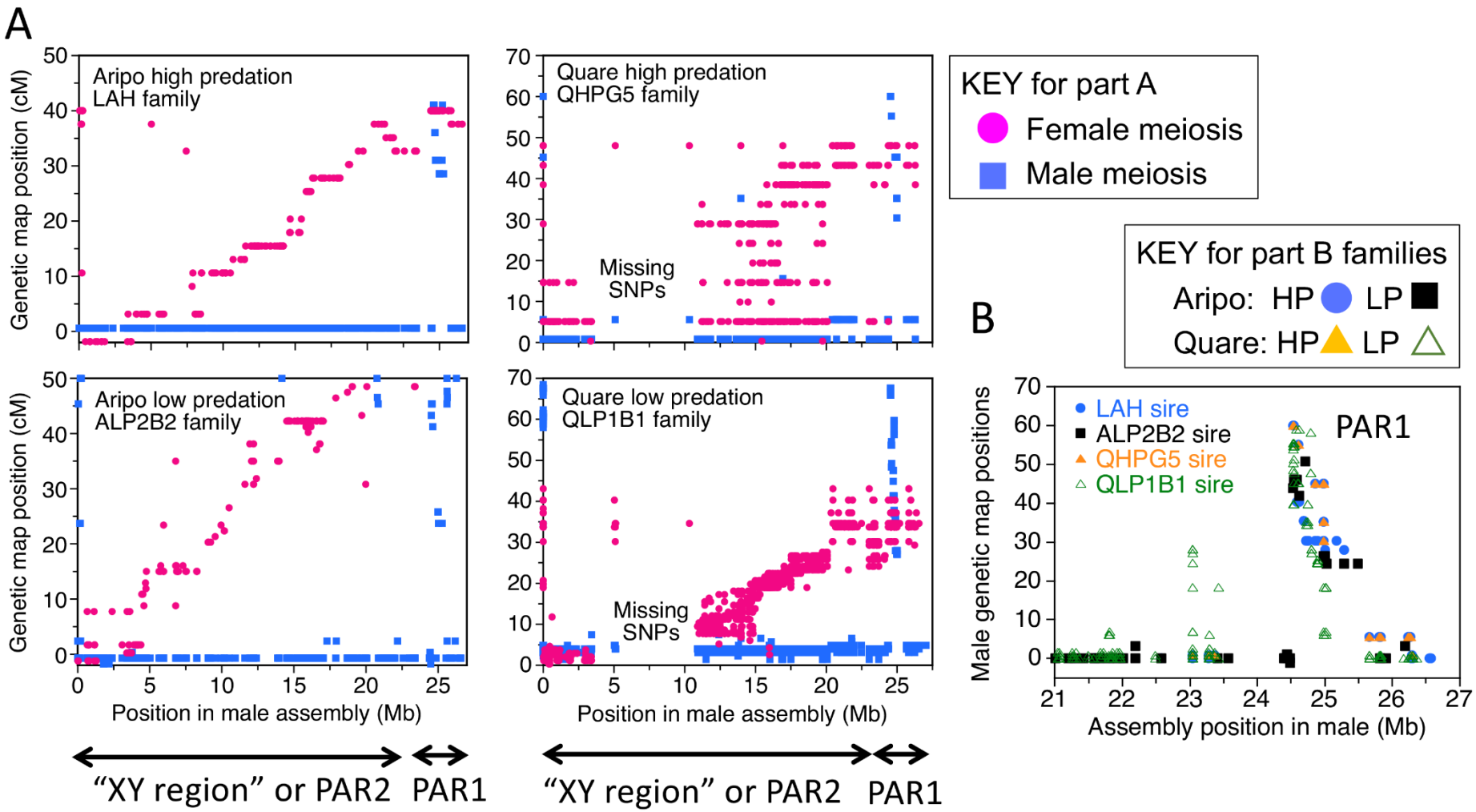

Figure 2
246x139 mm (x DPI)

Figure 3

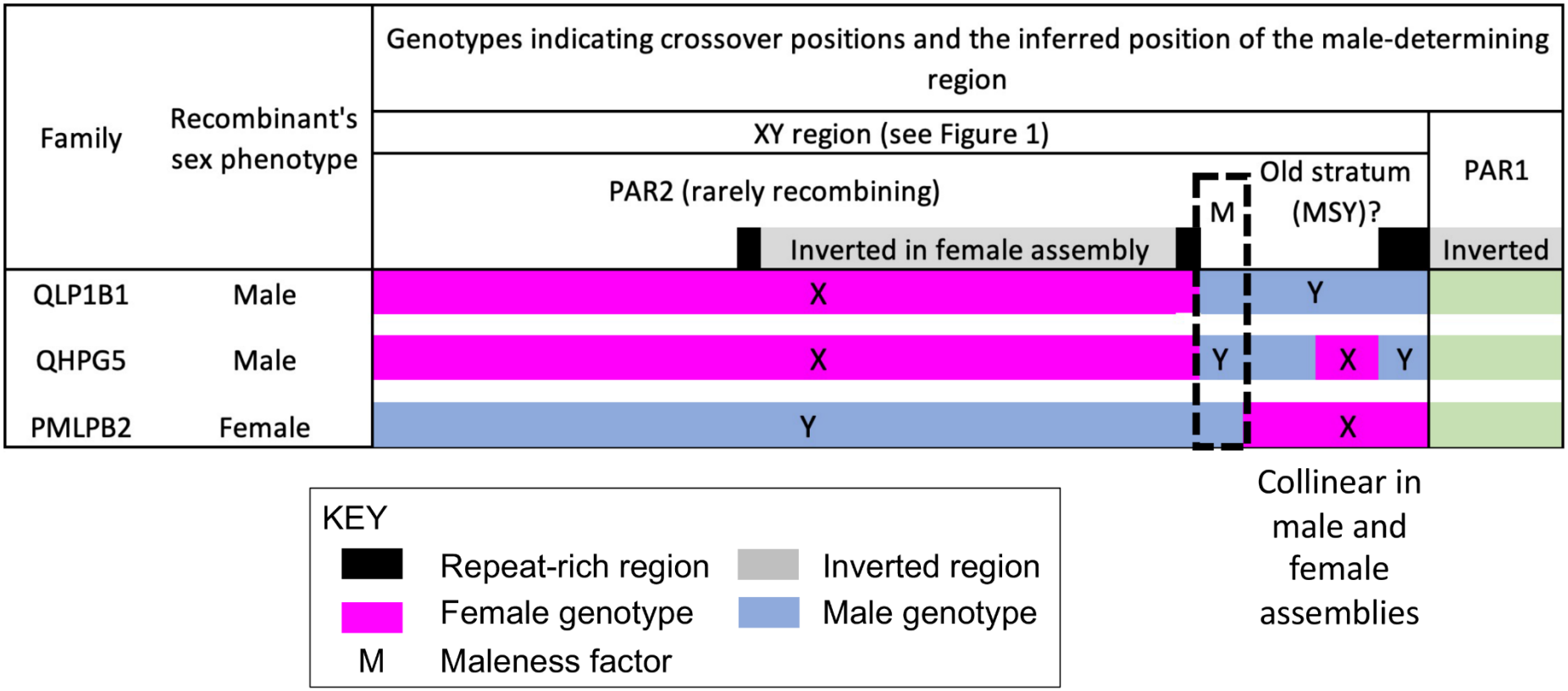

Figure 3
246x139 mm (x DPI)