GENE DUPLICATION TO THE Y CHROMOSOME IN TRINIDADIAN GUPPIES

Yuying Lin¹, Iulia Darolti¹, Benjamin L. S. Furman¹, Pedro Almeida², Benjamin A. Sandkam¹, Felix Breden³, Alison E. Wright⁴, Judith E. Mank¹,⁵

¹ Department of Zoology and Biodiversity Research Centre, University of British Columbia, Canada
² Department of Genetics, Evolution and Environment, University College London, United Kingdom
³ Department of Biological Sciences, Simon Fraser University, Canada
⁴ Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield
⁵ Biosciences, University of Exeter, Penryn Campus, United Kingdom

*Corresponding author E-mail: linyuying@zoology.ubc.ca
ABSTRACT

Differences in allele frequencies at autosomal genes between males and females in a population can result from two scenarios. Unresolved sexual conflict over survival produces allelic differentiation between the sexes. However, given the substantial mortality costs required to produce allelic differences between males and females at each generation, it remains unclear how many loci within the genome experience significant sexual conflict over survival. Alternatively, recent studies have shown that similarity between autosomal and Y sequence, arising from duplication onto the Y, can create perceived allelic differences, and this represents potentially resolved sexual conflict. However, Y duplications are most likely in species with large non-recombining regions, in part because they simply represent larger targets for duplications. We assessed the genomes of 120 wild-caught guppies, which experience extensive predation- and pathogen-induced mortality and have a relatively small ancestral Y chromosome. We identified seven autosomal genes that show allelic differences between male and female adults. Five of these genes show clear evidence of whole or partial gene duplication to the Y chromosome, suggesting that the male-specific region of the guppy Y chromosome, although relatively small, may nonetheless act as a hotspot for the resolution of sexual conflict. The remaining two genes show evidence of partial homology to the Y. Overall, our findings suggest that the guppy genome experiences a very low level of unresolved sexual conflict over survival, and instead the Y chromosome, despite its small ancestral size and recent origin, acts as a major mechanism of conflict resolution.

Keywords: gene duplications, Y chromosome, sexual conflict, Poecilia reticulata
Introduction

Differences in allele frequency between males and females at autosomal loci can result from two alternative sources. Numerous recent studies have used allelic differences between males and females within a population ($F_{ST}$, $D_{xy}$, etc) as a way to infer sex-differences in viability or survival, and therefore sexual conflict over mortality (Cheng & Kirkpatrick, 2016; Dutoit et al., 2018; Flanagan & Jones, 2017; Kasimatis et al., 2021; Lucotte et al., 2016; Wright et al., 2018, 2019). This approach assumes that allele frequencies are the same in males and females at conception, but diverge over the course of a generation for loci with alleles that benefit the survival of one sex at some survival cost to the other (intra-locus sexual conflict). We expect that sexual conflict over survival would produce a signature of allelic differentiation between the sexes as well as balancing selection (Mank, 2017). The latter results from all forms of intra-locus sexual conflict, not just that over survival, as alleles are selected for or against depending on whether they are present in males or females (Barson et al. 2015; Connallon & Clark 2014; Foerster et al. 2007; Hawkes et al. 2016; Johnston et al. 2013; Lonn et al. 2017; Ruzicka et al. 2019). Balancing selection resulting from sexual conflict may be a major factor in the maintenance of genetic polymorphisms within populations (Connallon & Clark 2014), therefore, the proportion of loci subject to unresolved sexual conflict within the genome, as well as the types of loci affected, may have important implications for a range of evolutionary factors, such as the speed and genetic basis of adaptation.

Intra-locus conflict over survival implies a significant mortality cost each generation. Modelling and simulation methods suggest that the sex-specific mortality rates necessary to generate significant allelic differences between the sexes within each generation are quite high for any one locus (Bissegger et al. 2019; Kasimatis, Ralph, & Phillips 2019; Ruzicka et al. 2020). This precludes the presence of large numbers of sites subject to sexual conflict due to survival, as the associated mortality load would simply be too great. Moreover, recent work has highlighted the potential that many perceived allelic sex differences between the sexes actually are the result of sequence homology between autosomal and sex-linked loci (Bissegger et al. 2019; Kasimatis et al. 2019; Mank, Shu, & Wright 2020; Ruzicka et al. 2020). The Y chromosome may preferentially retain duplicates that play an important role in male development or fitness.
(Bachtrog, 2013; Carvalho et al., 2015), and the process of Y duplication ironically offers a route to sexual conflict resolution even though it may lead to the mistaken perception of unresolved conflict acting on autosomal loci.

It therefore remains unclear how common and pervasive sexual conflict over survival is in natural populations, or what types of loci, if any, it is expected to target. Here we assess the potential for intersexual F\textsubscript{ST} based on a sample of 120 wild-caught adult guppies, which are expected to experience substantial predation- and pathogen-induced mortality. We identified seven genes that show evidence of allelic differences between adult males and females. Of those seven genes, five show evidence of Y duplicates, and two have sex-specific functions, including germline differentiation and sex hormone signaling. This is surprising, as although the size of the guppy Y varies substantially across populations and related species (Wright et al. 2017; Darolti et al. 2019; Ameida et al. 2021), the conserved, ancestral part of the Y is very small, spanning at most 5 Mb (Almeida et al., 2020; Darolti et al., 2019; Fraser et al., 2020). Given this small size, we might expect that the Y chromosome in guppies does not represent a large enough genomic target for sexual conflict resolution through gene duplication. Instead, our results show that the guppy Y, although a very small region of the genome, may indeed act as a hotspot for the resolution of conflict.

**Results**

We individually sequenced 60 male and 60 female wild-caught adults from three rivers in Trinidad to an average of ~30X coverage in males and ~20X coverage in females after filtering, trimming and quality control (see Materials and Methods for further details). The greater read depth in males was designed to aid detection of the Y chromosome, which is present in only one copy in each genome. Because the size of the non-recombining region of the sex chromosome varies across guppy populations (Almeida et al., 2020; Wright et al., 2017), and sex chromosome divergence results in allelic differences on the X and Y between males and females, we confined all our analyses to autosomal sequence, although we present the X chromosome (Chromosome 12) in figures for comparison.
From our sequencing data, we obtained 7,889,657 biallelic high quality filtered autosomal SNPs. We further filtered these to 253,375 SNPs present in annotated autosomal coding sequence. We focused on coding sequence for three reasons. First, non-coding sequence is enriched for repetitive elements, and the accumulation of repetitive elements on the ancestral guppy Y chromosome (Almeida et al., 2020) would lead to male-specific SNPs in repetitive elements that could bias our estimates of Y duplications. Second, previous work has suggested that unresolved sexual conflict on the autosome occurs primarily in coding sequence (Ruzicka et al., 2019), and this fits with the substantial evidence that sexual conflict over gene regulation can be relatively quickly resolved through sex-specific gene regulation (Kopp, Duncan, & Carroll 2000; Mank 2017; Wright et al. 2019). Finally, by focusing on functional coding sequence, we are able to determine the potential functional role of genes that are subject to sexual conflict.

**Intersexual F\textsubscript{ST}**

Given the low level of allelic divergence expected between males and females, it is critical to minimize false positives (Kasimatis et al. 2019). We therefore used three independent methods to estimate SNPs with elevated intersexual F\textsubscript{ST}. We identified SNPs that were 1) in the top 1% of the autosomal F\textsubscript{ST} distribution 2) were significant after permutation testing of samples (1000 replicates, P < 0.001) and 3) showed significant differences in male and female allele frequency based on Fisher’s exact test (P < 0.001) (Supplementary Fig. 1). We identified 504 autosomal coding sequence SNPs that were significant by all three of these measures (Supplementary Figs. 1 and 2), designated as high intersexual F\textsubscript{ST} SNPs.

We then identified autosomal coding genes with ≥ 3 high intersexual F\textsubscript{ST} SNPs (Bissegger et al. 2019; Tobler, Nolte, & Schlötterer 2017), resulting in seven loci with significant sexual differentiation. For each of these genes, average intersexual F\textsubscript{ST} was significantly greater than the autosomal average (Table 1, Supplementary Fig. 3).

Estimates of intersexual F\textsubscript{ST} and Tajima’s D can be biased due to relatedness of individuals within groups. Although our sampling design was balanced, with 10 females and 10 males collected at two separate sites for each of three rivers across the island of Trinidad (see
Almeida et al. 2020 for details), we could not account for relatedness in wild-caught samples at the time of collection. We therefore assessed pairwise kinship coefficients among all our samples using both KING (Manichaikul et al., 2010) and NgsRelate (Korneliussen & Moltke 2015), as implemented in ANGSD (Korneliussen, Albrechtsen, & Nielsen 2014). Neither method identified greater relatedness among male compared to female samples (Supplementary Fig. 4).

**Y duplications**

When mapping whole-genome data to a female reference genome, sequence similarity between the male-specific Y chromosome and the autosomes can lead to the perception of allelic differences between males and females. Recent work has shown that many genes with allelic sex differences are in fact autosomal loci that either have recent Y duplicates (Bissegger et al. 2019; Mank et al. 2020) or otherwise display sequence homology to the Y (Kasimatis et al., 2021). We can use differences in M:F read depth to identify these genes (Hall et al., 2013). Duplication of the complete coding sequence for an autosomal gene on the Y chromosome would produce an average M:F read depth of 1.5 (three copies in males, two in females), and subsequent tandem duplications on the Y would result in M:F read depth > 1.5. Partial Y duplications, or a full duplication followed by significant differentiation, would result in average M:F read depth > 1 and < 1.5 when mapping to a female genome after correcting for differences in library size.

Five of our seven sexually differentiated genes showed average M:F read depth significantly > 1 (Table 1, Supplemental Fig. 5) consistent with at least a partial duplication of the coding sequence on the Y chromosome (Fig. 1). Four of these Y-duplicated genes have annotated functions, two of which (Olr1496-like and Fezf1) have a known role in sexual differentiation (Table 1), as is expected for genes on the Y chromosome.

In order to differentiate whether the pattern of elevated M:F $F_{ST}$ is due to limited sequence homology between the autosomal genes and the Y chromosome, or results from at least partial gene duplication, we mapped the pattern of M:F read depth across each of these 5 genes (Fig. 1). The high M:F read depth for Olr1492-like, consistent across the full length of the single exon
of this gene, suggests a complete duplication, possibly in two copies (Fig. 1A). Similarly, the M:F read depth ratio for si:rp71-17i16.5 (Fig. 1C), Bcl11a (Fig. 1E), and Fezf1 (Fig. 1G) suggests a single or double copy duplication, of just the exons, possibly through retrotransposition. The high M:F read depth in ENSPRG05754 (Fig. 1I) suggests a partial Y duplication of the last exon of the gene, likely present in several Y copies.

Given sufficient evolutionary time, Y duplications will also accumulate male-specific SNPs (Tobler et al. 2017), leading to a signal of elevated M:F SNP density. Of the five sexually differentiated genes with greater read depth in males, none showed significantly higher M:F SNP density (Table 1), although all values were >1, suggesting that duplication to the Y is relatively recent, and consistent with the recent origin of this chromosome (Darolti et al. 2019).

Across our seven sexually differentiated genes, we did not observe evidence of duplications to the X chromosome, which would result in three copies in males and four in females and therefore an average M:F read depth significantly <1.

Sexual conflict over survival for autosomal genes

We identified just two sexually differentiated genes without significantly higher M:F read depth (Table 1), both of which displayed relatively modest levels of intersexual FST (0.0237 and 0.0098). At first consideration, this small number of loci with modest FST is consistent with limited unresolved sexual conflict and the expectation that large numbers of high intersexual FST loci would require unsustainable levels of sex-specific mortality each generation (Bissegger et al. 2019; Kasimatis et al. 2019; Ruzicka et al. 2020).

Genes subject to ongoing sexual conflict are expected to experience balancing selection, which is often measured with Tajima’s D (Bissegger et al. 2019; Cheng & Kirkpatrick 2016; Mank 2017; Wright et al. 2018). We therefore compared Tajima’s D for these two genes to the autosomal average, genes with known immune function, which are known to be subject to high levels of balancing selection (Andrés et al. 2009; Ferrer-Admetlla et al. 2008; Van Oosterhout 2009; Weedall & Conway 2010), and to other genomic categories (Fig. 2). Contrary to our expectation of elevated Tajima’s D for these two sexually differentiated genes, neither showed a Tajima’s D
value significantly greater than the autosomal average, and Tajima’s D for ENSPREG15023 is
significantly lower than the autosomal average (Fig. 2).

Many things can influence Tajima’s D estimates, and the lack of an elevated signature of
balancing selection relative to the remainder of the genome is not necessarily indicative of a
lack of sexual conflict. In order to understand the dynamics of these two genes in more detail,
we mapped M:F read depth as a function of genomic location and intersexual F_{ST} (Fig. 3).
Indeed, we found evidence of elevated M:F read depth for ENSPREG15023 across most of the
exon, consistent with either partial duplication or full exonic duplication to the Y followed by
partial degeneration. Many of the high F_{ST} SNPs in the gene have a high M:F read depth,
consistent with Y duplication or significant Y homology. The pattern for ENSPREG12439 is more
consistent with significant Y homology given the highly variable M:F read depth across the
gene. Importantly, all SNPs with high M:F F_{ST} values for this gene also show high M:F read
deepth. These results, coupled with the lack of elevated Tajima’s D, suggest that these two genes
also exhibit significant homology to the Y, and that this is driving the elevated intersexual F_{ST}
instead of unresolved sexual conflict over survival.

**Discussion**

We identified 504 coding sequence SNPs in the guppy autosomal genome which showed
significant differences in allele frequency between males and females. We used these to
identify seven autosomal genes with significant average intersexual F_{ST}. Our approach, based on
the intersection of three statistical methods, reduces the likelihood of false positives and
results in a high confidence gene list of intersexual F_{ST} (Table 1, Supplementary Figs. 1, 2, and
3). This list of genes can be used to assess the relative role of Y duplication in resolving conflict
and driving intersexual F_{ST}, versus the role of sex differences in mortality in sexual conflict of
autosomal genes.

**The Y chromosome in guppies as a locus for conflict resolution.**

Five of the seven high-confidence sexually differentiated genes showed evidence of Y
duplications based on elevated male-to-female read depth ratios (Table 1, Supplementary Fig.
Two of these (Olr1496-like and Fezf1) have sex-specific functions based on Gene Ontology and Gene Cards annotations, consistent with theory and empirical studies showing that Y chromosomes can accumulate gene duplicates with male-specific functions (Carvalho et al. 2015; Mahajan & Bachrog 2017). Specifically, Olr1496 (Olfactory factor 1496-like gene) plays a role in multiple aspects of germline development in *C. elegans* (Cho, Rogers, & Fay 2007), and Fezf1 (FEZ Family Zink Finger 1) plays an important role in the migration of gonadotropin-releasing neurons (Damla Kotan et al., 2014).

Y chromosomes in general represent a unique genomic environment. Although Y chromosome gene content degrades quickly following recombination suppression (Bachtrog, 2013), several recent studies indicate that older, large Y chromosomes, such as that in *Drosophila* (Carvalho et al. 2015; Koerich et al. 2008; Tobler et al. 2017), sticklebacks (Bissegger et al. 2019) and humans (Kasimatis et al. 2020), contain substantial numbers of genes duplicated from autosomes. Duplications may represent an important mechanism of sex chromosome divergence, and the Y chromosome may preferentially retain duplicates that play an important role in male development or fitness (Bachtrog, 2013; Carvalho et al., 2015), offering a hotspot for sexual conflict resolution within the genome.

However, it remains unclear how common Y duplications of autosomal genes occur in younger systems with a far smaller male-specific region. Although the size varies by population, the conserved non-recombining region of the Y chromosome in guppies is relatively small, spanning at most 5 Mb (Almeida et al., 2020; Darolti et al., 2019). Others have failed to recover evidence for even this limited region of Y degeneration (Charlesworth et al., 2020; Fraser et al., 2020). Our findings add further support for a region of recombination suppression across populations on Trinidad, as only a male-specific region of the Y can explain the M:F read depth differences we observe. Moreover, our work suggests that the guppy Y chromosome is dynamic with regard to gene content, and acts as a hotspot for gene duplications with male-specific functions despite its recent origin and small size. It is also possible that Y genes have duplicated to the autosomes, which would also produce a pattern of increased male read depth, although this is arguably less likely. Taken together, our work suggests that even homomorphic sex chromosomes may act as a hotspot of sexual conflict resolution. Moreover, our results further
emphasize the importance of accounting for Y gene duplications in scans for M:F $F_{ST}$, as the all of our sexually differentiated genes show evidence of Y duplication.

The Y-duplicated genes we identify here are not present in our previous list of Y-linked genes based on male-specific sequence (Almeida et al., 2020), or in other similar analyses (Fraser et al., 2020). This is not surprising, as duplications, particularly if recent, will still retain substantial homology to the autosomal copy and will not be detected when bioinformatically identifying sequence that is unique to male genomes. Consistent with recent duplications and limited divergence, none of our Y-duplicated genes exhibit significantly elevated average M:F SNP density (Table 1, Fig. 2), although the values are all >1.

We have previously (Wright et al., 2018) noted evidence of intersexual $F_{ST}$ in guppies (*Poecilia reticulata*). This finding was curious given that the data from this earlier work derived from a lab population, free of most of the pathogens and all the predators that would exacerbate sex-differences in mortality and predation (Wright et al., 2018). It was not clear how much of this signal, if any, was due to Y duplications. However, it is worth noting that elevated intersexual $F_{ST}$ was highest for genes with male-biased expression, as would be expected for genes with Y duplicates. We also did not observe a concomitant pattern of elevated Tajima’s D for these genes, which is inconsistent with sexual conflict over survival.

**Sexual conflict over survival targets few genes in the guppy autosomal genome**

Intra-locus sexual selection over survival or viability leads to allele frequency differences between the sexes over the course of a generation, as an allele increases the survival of one sex at a mortality cost to the other (Kasimatis et al. 2019; Mank 2017; Wright et al. 2018). The significant mortality costs required each generation to generate allele frequency differences between the sexes preclude large numbers of genes with significant M:F $F_{ST}$ (Bissegger et al. 2019; Kasimatis et al. 2019; Ruzicka et al. 2020). At most, we might expect a limited number of loci with significant allelic differentiation between males and females, and this would be most evident in wild species where males and females experience differences in predation, parasite or pathogen loads. Despite the fact that we sampled wild individuals of a classic prey species with very high potential for sexual conflict over survival, we observe just two loci in the genome
that show evidence of sexual conflict over survival based on M:F \( F_{ST} \). Both of these loci exhibit very low M:F \( F_{ST} \).

We expect that if these genes are indeed subject to sexual conflict over mortality, we would observe elevated Tajima’s D (Fig. 2), however this was not the case for either locus. Moreover, both of these loci exhibit patterns of M:F read depth consistent with significant Y homology (Fig. 3). This suggests that the potential for sexual conflict over survival is quite low, even in a species where we most expect to observe it.

Concluding remarks

Here we use a high-stringency filtering method to detect genes within the guppy genome that exhibit population genetic signatures expected from sexual conflict over survival. Although wild guppies are expected to have high potential for sexual conflict over survival, we in fact found no genes within the genome with patterns consistent with this. Instead, despite the small size of the conserved non-recombining region of the guppy Y, we observe five loci that show patterns consistent with autosome-to-Y duplications, and two more that are suggestive of Y homology of without duplication. This highlights the potential of even young, small Y chromosomes as regions of conflict resolution.

Materials and Methods

Data Collection and Genotyping

Samples were collected from three rivers, Aripo, Yarra, and Quare, in Trinidad in December 2016, in accordance with national collecting guidelines. In total, 10 males and 10 females were collected from one high predation and one low predation population in each river, resulting in 120 samples, which were sequenced individually with Illumina HISEQX. Further sequencing details are available in Almeida et al. (2020).

We used FastQC v0.11 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic 0.36 (Bolger et al., 2014) to remove adapter sequences and low-quality reads.
After quality control, we recovered ~30X average sequencing depth for males and ~20X sequencing depth for females. High quality reads were aligned against the *Poecilia reticulata* female reference genome (Ensembl GCA_000633615.2) (Künstner et al., 2016), using BWA 0.7.15 MEM algorithm (Li & Durbin 2009) with default parameters. We filled in mate coordinates and mate related flags, sorted alignment by coordinates, and marked PCR duplications with SAMtools-1.9 (Li et al., 2009).

We called genotypes across all the samples using the `mpileup` function from bcftools with the following parameters: --min-MQ 20 --min-BQ 20 --skip-indels -a FORMAT/AD, FORMAT/DP. After genotyping, we used VCFtools v0.1.16 (Danecek et al., 2011) to exclude SNPs that had either: (1) genotype quality < 25; (2) sequencing depth <0.5x or >10x of average depth; (3) missing data in > 10% of individuals; (4) minor allele frequency < 0.05 or (5) relative read depth between reference allele and alternative allele < 0.2 or > 0.8. In total, the autosomal filtered SNP dataset consisted of 7,889,657 biallelic SNPs. We extracted 253,375 SNPs in annotated coding sequences (Ensembl build Guppy Female 1.0) for downstream analysis. Finally, we confined our analysis to autosomal genes, but included the X chromosome (Chromosome 12) in figures as a means of comparison.

**Intersexual F\textsubscript{ST}**

In order to estimate intersexual allele frequency differences, we implemented Weir & Cockerham’s estimator of F\textsubscript{ST} (Weir & Cockerham 1984) between males and females using VCFtools v0.1.16 for each SNP in genome-wide coding sequence regions. We employed three methods jointly to identify SNPs exhibiting high F\textsubscript{ST}. First, we used a cut-off method, retaining SNPs in only the top 1% of autosomal F\textsubscript{ST} values. Second, we performed permutation tests by randomly assigning individuals to one of two sex groups to generate a null distribution of F\textsubscript{ST} across the genome. We determined significance for each SNP from 1000 replicates, using a P < 0.001 threshold. Finally, we performed Fisher’s exact test on SNPs to determine significance of allele frequency differences between males and females (P < 0.001). We denoted SNPs that were significant in all three of these measures as high F\textsubscript{ST} SNPs.
Using approaches to further limit false positives (Bissegger et al., 2019; Tobler et al., 2017), we identified 7 genes with $\geq$ 3 high intersexual F\textsubscript{ST} SNPs, which we designated as sexually differentiated genes. We calculated average intersexual F\textsubscript{ST} for all genes using VCFtools v0.1.16, respectively. We used Wilcoxon rank-sum test to indicate statistical difference in intersexual F\textsubscript{ST} between autosomal genes and other gene categories (sexually differentiated genes and genes on the sex chromosome).

**Relatedness Inference**

In order to avoid biases in calculating intersexual allele frequency differences due to relatedness among individuals, we used KING 2.2.7 (Manichaikul et al., 2010) to infer the pairwise degree of relatedness between individuals from estimated kinship coefficients. We first converted genotype data from the raw, unfiltered SNPs dataset to plink binary format using PLINK 1.9 (Purcell et al., 2007). In order to avoid potential biases from KING software (Ramstetter et al., 2017) and validations, we also used NgsRelate v2 (Korneliussen & Moltke 2015), implemented in ANGSD (Korneliussen et al. 2014) to infer genetic relatedness coefficients for each pair of individuals.

**Assessing Y Duplications of Sexually Differentiated Genes**

When using a female reference genome, reads from genes which have duplicated to the male-specific region of the Y chromosome will map back to original autosomal or X chromosome regions, resulting in elevated M:F coverage (Bissegger et al. 2019; Mank et al. 2020). For example, if an autosomal gene has one Y duplication, we would expect three copies in males (two autosomal and one Y-linked) and two copies in females, and therefore an average M:F read depth of 1.5. We first calculated M:F read depth for each coding sequence SNP from genotype data, as male coverage/female coverage, correcting for differences in average genomic read depth between males and females. In order to validate the M:F read depth ratio of sexually differentiated genes, we calculated male and female normalized read depth for 7 genes for each base pair, based on male and female pooled reads.

Genes on the Y chromosome will accumulate male-specific mutations over time, leading to an increased number of male-specific mutations as well as elevated M:F SNP density (Bissegger et
al. 2019; Mank et al. 2020). M:F SNP density for each gene was calculated as number of male SNPs/number of female SNPs.

**Tajima’s D**

Based on the filtered genotype data, we calculated Tajima’s D for the coding sequence of all autosomal genes using VCFtools v0.1.16. We compared mean Tajima’s D for autosomal genes (excluding those with immune function), genes with immune function (defined following Wright et al. 2017, Wright et al. 2018), and our seven sexually differentiated genes.

**Functional Annotations**

The small number of sexually differentiated genes precludes Gene Ontology enrichment analysis. We therefore cataloged functional annotations from the *Danio rerio* Gene Ontology (Ashburner et al., 2000; “The Gene Ontology resource: enriching a GOld mine,” 2021) and Gene Cards (Stelzer et al., 2016).

**Acknowledgements**

This was funded by a Chinese Scholarship Council Doctoral Scholarship to YL (No. 201906040216), a grant from the European Research Council (grant agreement 680951) and a Canada 150 Research Chair to JEM. We thank The Trinidad Ministry of Agriculture, Land and Fisheries and Indar Ramnarine for field permits. We thank XXX for helpful comments and suggestions on the manuscript.

**Data Accessibility Statement**

DNA-seq data are publicly available in the NCBI SRA (BioProject ID PRJEB39998). Scripts of all the analysis are available on GitHub XXXXX.

**Author Contributions**
J.E.M. and Y.L. designed the study. B.S., F.B. and J.E.M. collected samples. P.A. collected raw sequencing data. Y.L., I.D., B.L.S.F., P.A., B.S., A.E.W. and J.E.M. performed the data analysis.

J.E.M., Y.L., I.D., B.L.S.F., B.S., F.B. and A.E.W. contributed to the writing of the manuscript.

References

Almeida, P., Sandkam, B. A., Morris, J., Darolti, I., Breden, F., & Mank, J. E. (2020). Divergence and Remarkable Diversity of the Y Chromosome in Guppies. *Molecular Biology and Evolution*, 1–15. https://doi.org/10.1093/molbev/msaa257

Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., Gutenkunst, R. N., White, T. J., Green, E. D., Bustamante, C. D., Clark, A. G., & Nielsen, R. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msp190

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. In *Nature Genetics*. https://doi.org/10.1038/75556

Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics, 14*(2), 113–124. https://doi.org/10.1038/nrg3366

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., ... Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature, 528*(7582), 405–408. https://doi.org/10.1038/nature16062

Bissegger, M., Laurentino, T. G., Roesti, M., & Berner, D. (2019). Widespread intersex differentiation across the stickleback genome – The signature of sexually antagonistic selection? *Molecular Ecology, July*, 1–10. https://doi.org/10.1111/mec.15255

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carvalho, A. B., Vicoso, B., Russo, C. A. M., Swenor, B., & Clark, A. G. (2015). Birth of a new gene on the Y chromosome of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1516543112

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*. https://doi.org/10.1534/g3.120.401573

Cheng, C., & Kirkpatrick, M. (2016). Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1006170

Cho, S., Rogers, K. W., & Fay, D. S. S. (2007). The C. elegans Glycopeptide Hormone Receptor
Ortholog, FSHR-1, Regulates Germline Differentiation and Survival. *Current Biology.*
https://doi.org/10.1016/j.cub.2006.12.027

Connallon, T., & Clark, A. G. (2014). Balancing selection in species with separate sexes: Insights from fisher’s geometric model. *Genetics.* https://doi.org/10.1534/genetics.114.165605

Damla Kotan, L., Ian Hutchins, B., Ozkan, Y., Demirel, F., Stoner, H., Cheng, P. J., Esen, I., Gurbuz, F., Kenan Bicakci, Y., Mengen, E., Yuksel, B., Wray, S., & Kemal Topaloglu, A. (2014). Mutations in FEZF1 cause kallmann syndrome. *American Journal of Human Genetics.* https://doi.org/10.1016/j.ajhg.2014.08.006

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics.* https://doi.org/10.1093/bioinformatics/btr330

Darolti, I., Wright, A. E., Sandkam, B. A., Morris, J., Bloch, N. I., Farré, M., Fuller, R. C., Bourne, G. R., Larkin, D. M., Breden, F., & 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(38), 19031–19036. https://doi.org/10.1073/pnas.1905298116

Dutoit, L., Mugal, C. F., Bolívar, P., Wang, M., Nadachowska-Brzyska, K., Smeds, L., Yazdi, H. P., Gustafsson, L., & Ellegren, H. (2018). Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (Ficedula albicollis) genome. *Molecular Ecology.* https://doi.org/10.1111/mec.14789

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonet, T., Ramírez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J., & Casals, F. (2008). Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes. *The Journal of Immunology.* https://doi.org/10.4049/jimmunol.181.2.1315

Flanagan, S. P., & Jones, A. G. (2017). Genome-wide selection components analysis in a fish with male pregnancy. *Evolution.* https://doi.org/10.1111/evo.13173

Foerster, K., Coulson, T., Sheldon, B. C., Pemberton, J. M., Clutton-Brock, T. H., & Kruuk, L. E. B. (2007). Sexually antagonistic genetic variation for fitness in red deer. *Nature.* https://doi.org/10.1038/nature05912

Fraser, B. A., Whiting, J. R., Paris, J. R., Weadick, C. J., Parsons, P. J., Charlesworth, D., Bergero, R., Bemm, F., Hoffmann, M., Kottler, V. A., 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(10), 1789–1805. https://doi.org/10.1093/gbe/evaa187

Hall, A. B., Qi, Y., Timoshevskiy, V., Sharakhova, M. V., Sharakhov, I. V., & Tu, Z. (2013). Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. *BMC Genomics.* https://doi.org/10.1186/1471-2164-14-273

Hawkes, M. F., Gamble, C. E., Turner, E. C. R., Carey, M. R., Wedell, N., & Hosken, D. J. (2016). Intralocus sexual conflict and insecticide resistance. *Proceedings of the Royal Society B: Biological Sciences.* https://doi.org/10.1098/rspb.2016.1429

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature.* https://doi.org/10.1038/nature12489

Kasimatis, K. R., Abraham, A., Ralph, P. L., Kern, A. D., Capra, J. A., & Phillips, P. C. (2021).
Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics, 217*(1). https://doi.org/10.1093/genetics/iyaa015

Kasimatis, K. R., Ralph, P. L., & Phillips, P. C. (2019). Limits to genomic divergence under sexually antagonistic selection. *G3: Genes, Genomes, Genetics, 9*(11), 3813–3824.
https://doi.org/10.1534/g3.119.400711

Koerich, L. B., Wang, X., Clark, A. G., & Carvalho, A. B. (2008). Low conservation of gene content in the Drosophila Y chromosome. *Nature*. https://doi.org/10.1038/nature07463

Kopp, A., Duncan, I., & Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature*. https://doi.org/10.1038/35046017

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics, 15*(1), 1–13. https://doi.org/10.1186/s12859-014-0356-4

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv509

Künstner, A., Hoffmann, M., Fraser, B. A., Kottler, V. A., Sharma, E., Weigel, D., & Dreyer, C. (2016). The genome of the trinidadian guppy, Poecilia reticulata, and variation in the Guanapo population. *PLoS ONE, 11*(12), 1–25.
https://doi.org/10.1371/journal.pone.0169087

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp352

Lonn, E., Koskela, E., Mappes, T., Mokkonen, M., Sims, A. M., & Watts, P. C. (2017). Balancing selection maintains polymorphisms at neurogenetic loci in field experiments. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1621228114

Lucotte, E. A., Laurent, R., Heyer, E., Ségurel, L., & Toupance, B. (2016). Detection of allelic frequency differences between the sexes in humans: A signature of sexually antagonistic selection. *Genome Biology and Evolution, 8*(5), 1489–1500.
https://doi.org/10.1093/gbe/evw090

Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of y chromosome gene content in flies. *Nature Communications, 8*(1). https://doi.org/10.1038/s41467-017-00653-x

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq559

Mank, J. E. (2017). Population genetics of sexual conflict in the genomic era. *Nature Reviews Genetics, 18*(12), 721–730. https://doi.org/10.1038/nrg.2017.83

Mank, J. E., Shu, J. J., & Wright, A. E. (2020). Signature of sexual conflict is actually conflict resolved. In *Molecular ecology* (Vol. 29, Issue 2, pp. 215–217).
https://doi.org/10.1111/mec.15311

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-
genome association and population-based linkage analyses. *American Journal of Human Genetics*. https://doi.org/10.1086/519795

Ramstetter, M. D., Dyer, T. D., Lehman, D. M., Curran, J. E., Duggirala, R., Blangero, J., Mezey, J. G., & Williams, A. L. (2017). Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics*. https://doi.org/10.1534/genetics.117.1122

Ruzicka, F., Dutoit, L., Czuppon, P., Jordan, C. Y., Li, X., Olito, C., Runemark, A., Svensson, E. I., Yazdi, H. P., & Connallon, T. (2020). The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, 4(5), 398–415. https://doi.org/10.1002/evl3.192

Ruzicka, F., Hill, M. S., Pennell, T. M., Flis, I., Ingleby, F. C., Mott, R., Fowler, K., Morrow, E. H., & Reuter, M. (2019). Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. In *PLoS Biology* (Vol. 17, Issue 4). https://doi.org/10.1371/journal.pbio.3000244

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*. https://doi.org/10.1002/cpbi.5

The Gene Ontology resource: enriching a GOld mine. (2021). *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1113

Tobler, R., Nolte, V., & Schlötterer, C. (2017). High rate of translocation-based gene birth on the Drosophila Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), 11721–11726. https://doi.org/10.1073/pnas.1706502114

Van Oosterhout, C. (2009). A new theory of MHC evolution: Beyond selection on the immune genes. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2008.1299

Weedall, G. D., & Conway, D. J. (2010). Detecting signatures of balancing selection to identify targets of anti-parasite immunity. In *Trends in Parasitology*. https://doi.org/10.1016/j.pt.2010.04.002

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

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

Wright, A. E., Fumagalli, M., Cooney, C. R., Bloch, N. I., Vieira, F. G., Buechel, S. D., Kolm, N., & Mank, J. E. (2018). Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters*, 2(2), 52–61. https://doi.org/10.1002/evl3.39

Wright, A. E., Rogers, T. F., Fumagalli, M., Cooney, C. R., & Mank, J. E. (2019). Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict. *Molecular Ecology*. https://doi.org/10.1111/mec.15115
### Table 1. Summary statistics of sexually differentiated genes. Statistically significant values are shown in bold.

| ENSPREG\(^1\) (Gene Name if known) | Chrom | Mean \(F_{ST}\)\(^2\) | M:F read depth\(^3\) | M:F SNP density\(^3\) | Gene functions\(^4\) |
|---------------------------------|-------|-----------------|-----------------|-----------------|-----------------|
| 03921 (Olr1496-like)            | LG7   | 0.1645          | 1.3809          | 1.3000          | Immune response & germline differentiation |
| 03854 (si:rp71-17i16.5)         | LG7   | 0.0892          | 1.1580          | 1.1200          | Phosphatidylinositol 3-kinase signalling, involved in immune, inflammatory and allergic responses |
| 09757 (BCL11A)                  | LG15  | 0.0419          | 1.3280          | 1.0000          | Hematopoietic cell differentiation & brain development |
| 07572 (Fezf1)                   | LG6   | 0.0412          | 1.5082          | 1.0077          | Nervous system development, migration of gonadotropin-releasing hormone neurons |
| 05754                           | LG15  | 0.0238          | 1.8887          | 0.9445          | N/A |
| 15023                           | LG18  | 0.0237          | 1.0829          | 1.2143          | N/A |
| 12439                           | LG22  | 0.0098          | 1.0802          | 1.0417          | N/A |

1 Ensembl guppy gene number  
2 Significance based on Wilcoxon’s rank-sum test \((P < 0.01)\), compared to autosomal average.  
3 Significance based on Fisher’s Exact Test \((P < 0.05)\), relative to genome-wide M:F SNP density.  
4 From Gene Ontology (Ashburner et al., 2000; “The Gene Ontology resource: enriching a GOld mine,” 2021) and Gene Cards (Stelzer et al., 2016).
**Fig. 1.** Elevated M:F read depth ratio indicates partial gene duplication of Olr1496-like (Panels A and B), si:rp71-17i16.5 (Panels C and D), BCL11A (Panels E and F), Fezf1 (Panels G and H), ENSPRE05754 (Panels I and J). Left panels show M:F read depth as a function of physical position (in Mb), with coding regions shaded in grey. Panels on the right show intersexual F_{ST} and M:F read depth ratio for each SNP. 95% autosomal confidence intervals for M:F read depth are indicated by horizontal red lines.

**Fig. 2.** Distribution of Tajima’s D among different gene categories. Significant differences to autosomal distribution are indicated (Wilcoxon rank-sum test, \(*P < 0.01\), \(**P < 0.001\), \(***P < 0.0001\), ns: non-significant).
Fig. 3. Normalized M:F read depth ratio for ENSPREG15023 (panels A and B) and ENSPREG12439 (panels C and D). Left panels show M:F read depth as a function of physical position (in Mb), with coding regions shaded in grey. Panels on the right show intersexual $F_{ST}$ and M:F read depth ratio for each SNP. 95% autosomal confidence intervals for M:F read depth are indicated by horizontal red lines.
Supplementary Fig. 5. Normalized M:F read depth ratio, with seven sexually differentiated genes indicated in orange. Steel blue and grey: autosomes; Light grey: sex chromosome (Chromosome 12).