Guppy Y chromosome integrity maintained by incomplete recombination suppression

Iulia Darolti[1], Alison E. Wright[2], Judith E. Mank[1,3]

1 Department of Zoology, University of British Columbia, Canada

2 Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

3 Department of Genetics, Evolution and Environment, University College London, United Kingdom

Corresponding author: Iulia Darolti, e-mail: darolti@zoology.ubc.ca

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

The loss of recombination triggers divergence between the sex chromosomes and promotes degeneration of the sex-limited chromosome. Several livebearers within the genus *Poecilia* share a male-heterogametic sex chromosome system that is roughly 20 million years old, with extreme variation in the degree of Y chromosome divergence. In *P. picta*, the Y is highly degenerate and associated with complete X chromosome dosage compensation. In contrast, although recombination is restricted across almost the entire length of the sex chromosomes in *P. reticulata* and *P wingei*, divergence between the X and the Y chromosome is very low. This clade therefore offers a unique opportunity to study the forces that accelerate or hinder sex chromosome divergence. We used RNA-seq data from multiple families of both *P. reticulata* and *P. wingei*, the species with low levels of sex chromosome divergence, to differentiate X and Y coding sequence based on sex-limited SNP inheritance. Phylogenetic tree analyses reveal that occasional recombination has persisted between the sex chromosomes for much of their length, as X- and Y-linked sequences cluster by species instead of by gametolog. This incomplete recombination suppression maintains the extensive homomorphy observed in these systems. In addition, we see differences between the previously identified strata in the phylogenetic clustering of X-Y orthologs, with those that cluster by chromosome located in the older stratum, the region previously associated with the sex-determining locus. However, recombination arrest appears to have expanded throughout the sex chromosomes more gradually instead of through a stepwise process associated with inversions.

Keywords: gametologs, non-recombining region, sex-linked genes, poecilid

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

INTRODUCTION

A common feature of sex chromosomes observed across a diverse array of taxa is loss of recombination, which can ultimately culminate in extreme differences in size and gene content between the sex chromosomes (Bachtrog, et al. 2011; Bull 1983; Charlesworth, et al. 2005). Non-recombining regions experience a reduction in the efficiency of selection to remove deleterious mutations as a result of decreased effective population size and accentuated Hill-Robertson effects (Charlesworth and Charlesworth 2000; Charlesworth, et al. 2005). As a consequence, over time, initially identical sex chromosomes are expected to diverge from each other in gene content and nucleotide sequence as the sex-limited chromosome accumulates deleterious mutations and degenerates (Bachtrog 2013; Charlesworth and Charlesworth 2000).

It is clear, however, that the degree of sex chromosome divergence does not always correlate with age (Bachtrog, et al. 2014; Stock, et al. 2011), and therefore Y degeneration is not inevitable, nor is the rate predictable. Because of this, one of the persistent mysteries of sex chromosome evolution is why some X and Y chromosomes show extensive divergence from each other while other systems remain homomorphic (Bachtrog, et al. 2014).

Variation in the rate of Y chromosome divergence across taxa could be the effect of different processes leading to recombination suppression. Recombination arrest is thought to initially cover the region containing the sex-determining locus and subsequently expand over larger portions of the sex chromosomes (Charlesworth, et al. 2005). The expansion of the non-recombining region can occur in a stepwise manner, through successive recombination suppression events. This results in distinct regions with different levels of sequence divergence between the gametologs, referred to as evolutionary strata of divergence (Bachtrog 2013; Charlesworth, et al. 2005; Lahn and Page 1999). Intra-chromosomal rearrangements such as large-scale inversions could result in rapid Y chromosome decay. This mechanism of recombination suppression, often assumed to be the main driver of sex chromosome divergence, would instantaneously prevent recombination throughout the inverted region (Charlesworth, et al. 2005). Indeed, sex chromosome divergence in some older systems shows the expected signatures of strata formation via inversions, namely clusters of gametologs with similar divergence estimates

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

(Cortez, et al. 2014; Handley, et al. 2004; Lahn and Page 1999; Wright, et al. 2014; Wright, et al. 2012).

However, newly evolved systems show less support for the classic model of sex chromosome evolution (Chibalina and Filatov 2011; Muyle, et al. 2012; Roesti, et al. 2013). In many younger systems, sequence divergence and recombination suppression between the gametologs occurs heterogeneously across the length of the sex chromosomes, suggesting that recombination suppression is a more gradual process than would be expected from the fixation of inversions (Almeida, et al. 2019; Bergero, et al. 2013; Chibalina and Filatov 2011; Natri, et al. 2013; Nicolas, et al. 2004). Moreover, if recombination suppression occurs primarily via mechanisms other than inversions, infrequent X-Y recombination events could persist and prevent the sex-limited chromosome from degenerating (Stock, et al. 2011; Stock, et al. 2013). This permeability of recombination suppression can act to obscure the boundaries between strata, as well as between the non-recombining region and the PAR (Chibalina and Filatov 2011).

Previous studies have used RNA-seq data and analyses of SNP segregation patterns in families to isolate sex-linked genes from autosomal or pseudoautosomal genes and obtain gametologous (X and Y) sequences (Chibalina and Filatov 2011; Hough, et al. 2014; Martin, et al. 2019; Muyle, et al. 2017; Muyle, et al. 2018; Veltsos, et al. 2019). This approach makes it possible to test for the presence of evolutionary strata by estimating divergence between gametologs and identifying clusters of sex-linked genes with different divergence rates (Lahn and Page 1999; Ross, et al. 2005; Wright, et al. 2014). While these analyses can be used to date recombination suppression events and distinguish between the different evolutionary strata of heteromorphic sex chromosomes, they are also useful to define the boundaries between the PAR and the non-recombining regions of younger, less differentiated systems (Campos, et al. 2017).

The common guppy (*Poecilia reticulata*) and its sister species the Endler's guppy (*Poecilia wingei*) share the same male-heterogametic sex chromosome system (Darolti, et al. 2019; Morris, et al. 2018). Our previous analysis of coverage and polymorphism data in males and females revealed evidence of two candidate evolutionary strata in both species. Stratum I, likely predating the divergence of *P. reticulata* and *P. wingei* (Darolti, et al. 2019; Wright, et

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

al. 2017), corresponds to an area around the sex-determining locus where recombination in males has previously been undetectable (Winge 1927, 1922; Yamamoto 1975), and exhibits mild Y degeneration (Darolti, et al. 2019; Wright, et al. 2017). Stratum II varies both within and between these two species, and exhibits very rare X-Y recombination events (Winge 1927, 1922; Yamamoto 1975) with the vast majority of male recombination confined to the ends of the chromosome (Bergero, et al. 2019). Despite occasional recombination, we previously detected low levels of X-Y divergence in Stratum II (Darolti, et al. 2019; Wright, et al. 2017), suggesting that either recombinants are selected against, or that the rate of recombination is sufficiently low that it does not fully counter the accumulation of Y chromosome mutations.

Furthermore, although recombination suppression was initially assumed to be quite recent based on the observed low level of divergence between the X and Y chromosomes, we recently showed that this sex chromosome system is in fact far older than expected, as it predates the common ancestor with *Poecilia picta*, roughly 20 million years ago (Darolti, et al. 2019; Meredith, et al. 2011). Curiously, despite the low levels of X and Y differentiation in *P. reticulata* and *P. wingei* (Darolti, et al. 2019; Morris, et al. 2018; Wright, et al. 2017), in *P. picta* the Y is highly diverged from the X and demonstrates complete X chromosome dosage compensation in males (Darolti, et al. 2019). This clade with a shared sex chromosome system and a range of sex chromosome divergence presents a unique comparative opportunity to understand the forces that accelerate or retard Y divergence.

Here, we use the probability-based approach, SEX-DETector (Muyle, et al. 2016), to infer autosomal and sex-linked genes in *P. reticulata* and *P. wingei* based on RNA-seq data from families. We used phylogenetic and sequence divergence analyses of these sex-linked genes to further characterize the previously defined evolutionary strata and to identify the forces that prevent large-scale sex chromosome degeneration over long evolutionary time. We found significantly more sex-linked genes in *P. wingei* than in *P. reticulata*, consistent with the expansion of the non-recombining region in the former. The previously defined non-recombining regions are significantly enriched in sex-linked genes, and we found evidence of recombination suppression before the separation of these two species for four genes in Stratum I. However, the X-Y sequence divergence between genes in the two evolutionary strata is not significantly different, and the same is true between genes in the PAR versus

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

the non-recombining region, suggesting that recombination arrest has evolved gradually. A phylogenetic analysis of X- and Y-linked sequences reveals that the extensive homomorphism of the poeciliid sex chromosomes is maintained by incomplete recombination suppression. Taken together, our results show that although recombination is largely suppressed across the entire length of the X and Y chromosomes, rare recombination events maintain the integrity of Y coding sequence and expression by preventing large-scale degradation of Stratum II. Our results present an integrated view of how occasional recombination events can retard divergence of sex chromosomes and maintain homomorphy.

MATERIALS AND METHODS

Sample collection and sequencing

We established three *P. reticulata* and two *P. wingei* families, and sampled parents and F1 progeny. For generating each family, we used a male and a virgin female either from a *P. reticulata* outbred laboratory population descended from a high predation population of the Quare River, Trinidad (Kotrschal, et al. 2013), or from our *P. wingei* laboratory population, derived from a strain maintained by a UK fish hobbyist. We only sampled families where the number of F1 progeny included at least five male and five female individuals, as this is the minimum number of offspring per family required to reliably identify sex-linked genes using the SEX-DETector software (Muyle, et al. 2016).

From each individual, we collected the posterior region of the fish behind the anal fin, which we preserved in RNAlater at -20°C prior to RNA preparation. We extracted RNA from these samples using the RNeasy Kit (Qiagen), following the manufacturer's instructions with an on-column DNase treatment. Libraries were prepared at the SciLife Lab in Uppsala, Sweden, following standard protocol. RNA was sequenced on an Illumina HiSeq 2500 platform with 125 bp paired-end reads, resulting in an average of 43 million (*P. reticulata*) and 41 million (*P. wingei*) RNA-seq reads per sample. We assessed sample quality using FastQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by adaptor removal and trimming using Trimmomatic v0.36 (Lohse, et al. 2012). We trimmed regions with average Phred score < 15 in sliding windows of four bases, reads with Phred score < 3 for leading and trailing bases, as well as paired-end reads with either read pair shorter than 50 bp. Following

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

trimming, we had, on average per sample, 29 million reads for *P. reticulata* and 28 million reads for *P. wingei* (Supplementary Table 1).

Constructing and filtering *de novo* transcriptome assemblies

We pooled reads from all samples of each species (n = 36 for *P. reticulata*; n = 24 for *P. wingei*) to construct species-specific *de novo* transcriptome assemblies (Supplementary Table 2) using Trinity v2.5.1 (Grabherr, et al. 2011) with default parameters, as per previously implemented methods (Bloch, et al. 2018; Harrison, et al. 2015; Wright, et al. 2019b). We then filtered the assemblies to remove redundancy and non-coding RNA. First, we used the Trinity align_and_estimate_abundance.pl script which maps RNA-seq reads to the transcriptomes using Bowtie2 (Langmead, et al. 2009), suppressing unpaired and discordant alignments, and estimates transcript abundance for each sample using RSEM v1.2.25 (Li and Dewey 2011). We then selected the best isoform for each gene based on the highest average expression. In cases where multiple isoforms had the highest expression, we chose the longest isoform as the best isoform. We further filtered the assemblies for non-coding RNA by removing transcripts with a blast hit to *Poecilia formosa* (PoeFor_5.1.2) or *Oryzias latipes* (MEDAKA1) ncRNA sequences obtained from Ensembl 84 (Flicek, et al. 2014). Lastly, we used Transdecoder v5.2.0 (http://transdecoder.github.io) with default parameters to remove transcripts missing an open-reading frame and transcripts with open-reading frames shorter than 150bp.

Assigning chromosomal position

We downloaded *P. reticulata* genes from NCBI RefSeq (Guppy_female_1.0+MT, RefSeq assembly accession: GCF_000633615.1) and identified the longest isoform for each gene. We BLASTed the best isoform sequences against the filtered *P. reticulata* and *P. wingei* transcriptome assemblies using BLASTn v2.3.0 (Altschul, et al. 1990) with an e-value cutoff of $10e^{-10}$ and a 30% minimum percentage identity. For genes mapping to multiple *de novo* transcripts, we selected the top blast hit based on the highest BLAST bit-score, a measure of sequence similarity. We assigned positional information on *P. reticulata* chromosomes to *P. reticulata* and *P. wingei* transcripts, based on the chromosomal location of genes in the reference.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Previously (Darolti, et al. 2019), we have generated pairwise alignments between the *P. reticulata* reference genome and female *de novo* genome assemblies of *P. wingei* and *P. picta* using LASTZ v1.04 (Harris 2007) and the UCSC chains and nets pipeline (Kent, et al. 2003). We recovered a region of the sex chromosome that is inverted in *P. reticulata* relative to *P. wingei* and *P. picta* (Supplementary Figure 1). As this analysis was run on female data alone, we inferred this inverted region to be on the X chromosome rather than the Y. It is likely that this inversion has occurred once in *P. reticulata* instead of independently in *P. wingei* and *P. picta*, and is specific to the inbred strain on which the reference genome assembly was built. Considering this, here, we ensured to account for the coordinates of the discovered inversion when assigning *de novo* transcripts with positional information on the *P. reticulata* sex chromosomes.

Inferring autosomal and sex-linked genes

In order to identify sex-linked genes we used the probability-based method SEX-DETector (Muyle, et al. 2016), which analyses the genotypes of individuals in a cross (parents and male and female offspring) to infer the segregation mode of each contig. The model can distinguish between three segregation types: autosomal, sex-linked with both the X and Y copies present (X/Y pair) and sex-linked with the X copy present but the Y copy lost or lowly expressed (X-hemizygous mode). For identifying sex-linked genes with X and Y alleles, SEX-DETector requires gametologs to co-assemble in one single transcript instead of separate transcripts. Co-assembly makes it possible to identify X/Y SNPs and is necessary to differentiate Y-linked sequences from autosomal genes with male-limited expression. Thus, the SEX-DETector algorithm can more effectively detect X/Y pairs in sex chromosome systems that have low or intermediate levels of divergence. While X-hemizygous contigs in old systems can be identified using this method, some may in fact be X/Y pairs whose sequences were assembled into separate contigs due to high levels of divergence (Muyle, et al. 2016). To avoid assembly of X and Y copies of the same transcript into separate contigs and prevent wrongly inferring contigs as X-hemizygous, we used CAP3 (Huang and Madan 1999), with a 90% minimum percent similarity between sequences, to further assemble contigs. The final filtered transcriptome assemblies contained a total of 19,935 *P. reticulata* transcripts and 19,361 *P. wingei* transcripts (Supplementary Table 2). The resulting *P. reticulata* and *P. wingei*

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

assemblies are of equivalent quality (Supplementary Table 2), which gives us a similar power to detect sex-linked loci in these two species.

Based on the SEX-DETector pipeline, we mapped reads onto the filtered assemblies using BWA v0.7.12 (Li and Durbin 2009) and then merged and sorted all libraries of each family separately using SAMtools v1.3.1 (Li, et al. 2009). We genotyped individuals at each locus using reads2snp v2.0 (http://kimura.univ-montp2.fr/PopPhyl/), with a minimum number of three reads for calling a genotype (option -min 3), a minimum base quality of 20 (option -bqt 20), a minimum mapping quality of 10 (option -rqt 10), the -aeb option for allowing alleles to have different expression levels, which is important for sex chromosome analyses as the Y copy can be less expressed than the X copy, and the paraclean option disabled (option -par 0) to avoid removal of paralogous positions since X and Y copies can resemble paralogs. We then used SEX-DETector to infer within each family the segregation type for each transcript, using a minimum posterior segregation type probability of 0.8, and to obtain for each family X and Y sequences of each sex-linked gene.

All subsequent analyses were run on a sex-linked gene dataset for each species comprised of the pooled sex-linked loci across replicate families. In the case of genes identified as sex-linked in multiple families, we selected the X and Y sequence pairs that contained the highest number of SNP differences as the representative sex-linked sequences for that species. This has increased the power of our downstream analyses, in particular of the divergence estimates for X- and Y-linked gametologs.

Phylogenetic analysis

In addition to the *P. reticulata* and *P. wingei de novo* transcript sequences, we obtained *Oryzias latipes* (MEDAKA1), *Xiphophorus maculatus* (Xipmac4.4.2) and *Poecilia formosa* (PoeFor_5.1.2) transcripts from Ensembl 84 and identified the longest transcript for each gene. We determined orthology across all these species using a reciprocal BLASTn with an e-value cutoff of $10e^{-10}$ and a minimum percentage identity of 30%. We then used BLASTx to identify open reading frames in each orthologous group.

We conducted a phylogenetic analysis of X and Y-linked sequences together with their orthologs in outgroup species to investigate the history of recombination suppression on the

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

sex chromosomes. First, we aligned sequences using PRANK v140603 (Löytynoja and Goldman 2008) and removed gaps in the alignments. We generated maximum likelihood phylogenetic trees with RAxML v8.2.12 (Stamatakis 2014), using the rapid bootstrap algorithm with the GTRGAMMA model and 100 bootstraps. We used Geneious v2019.2.3 (Kearse, et al. 2012) to concatenate the alignment files of sex-linked loci in both *P. reticulata* and *P. wingei* into a single file. Using this concatenated file, we ran RAxML using the rapid bootstrap algorithm with the GTRGAMMA model and 100 bootstraps to obtain consensus phylogenetic trees. Additionally, we repeated the phylogenetic analysis by constructing maximum likelihood phylogenetic trees using the GTRGAMMA model and 100 bootstrap replications in MEGA X (Kumar, et al. 2018). We used FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) to plot all phylogenetic trees.

Rates of evolution of sex-linked genes

For each identified sex-linked gene, we estimated synonymous X-Y divergence (dS_{XY}) using the *yn00* program in PAML v4.8 (Yang 2007), following the Yang and Nielsen method (Yang and Nielsen 2000). Within each species, we tested for differences in dS_{XY} between different gene categories (genes on the PAR 0-15Mb, the entire non-recombining region 15-26Mb, Stratum I 21-26Mb or Stratum II 15-21Mb) and between the species using Wilcoxon rank sum tests in R (R Core Team 2015).

We also estimated rates of evolution of X and Y sequences based on the multiple sequence alignments described above. We first masked poorly aligned regions with SWAMP v31-03-14 (Harrison, et al. 2014). We then used branch model 2 from the CODEML package in PAML to estimate rates of nonsynonymous (d_N) and synonymous (d_S) substitutions for the X- and Y-linked branches in each tree. We used the inferred phylogenetic trees described above to generate an unrooted tree for each orthologous group, which is required in the CODEML analyses. We used 1,000 permutation test replicates to identify significant differences in d_N/d_S ratios between the X and Y sequences, between the gene categories and between the two species.

Expression of X- and Y-linked genes

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Within each species, we mapped reads from each male individual to the identified X- and Y-linked gene sequences and estimated transcript abundance for homologous X- and Y-linked genes using RSEM v1.2.25 (Li and Dewey 2011). For each individual, we calculated Y/X expression at each sex-linked gene and then obtained average Y/X expression ratio across all male individuals of that species.

RESULTS

We used the genotypes of parents and offspring from multiple *P. reticulata* and *P. wingei* families to infer sex-linkage, which includes genes expressed from both the X and Y chromosome (X/Y genes) and genes only expressed from the X (X0 genes), following the SEX-DETector pipeline (Table 1 and Table 2). By mapping inferred sex-linked genes to known *P. reticulata* transcripts and assigning them with chromosomal position, we were able to verify that the vast majority of them (83% in *P. reticulata* and 92% in *P. wingei*) are indeed found on the sex chromosome, *P. reticulata* chromosome 12 (Table 1 and Table 2), indicating our false-positive rate is quite low. Genes inferred to have a sex-linked inheritance pattern that mapped outside the sex chromosomes are distributed across multiple autosomes and a few unplaced scaffolds (Supplementary Table 3). We found no enrichment of sex-linked genes on any of the autosomes and there is also little overlap between the two species in the autosomes that contain genes with a sex-linked inheritance pattern (Supplementary Table 3).

For all subsequent analyses, we only used sex-linked genes which map to chromosome 12. Only three genes were inferred to be sex-linked in all three *P. reticulata* families, while 142 genes show signatures of sex-linkage in both *P. wingei* families. Given our overall low false positive rate, and the fact that the small clutch size of these species limits our power to detect sex-linkage in any one family, we pooled all loci identified as sex-linked across replicate families within each species for following analyses.

SEX-DETector inferred only two X0 genes in *P. reticulata* and none in *P. wingei*. Both *P. reticulata* X0 genes are located in the sex chromosome non-recombining region (one at 24.9 Mb in the previously identified Stratum I and the other at 17.2 Mb in Stratum II). Genetic diversity can affect our likelihood of identifying X-hemizygous genes, as these are identified by the presence of polymorphisms on the X copy. Thus, there is a lower probability of detecting X-hemizygous genes in *P. wingei* compared to *P. reticulata* as the sampled *P. wingei*

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

population is more inbred and has a lower genetic diversity than the sampled *P. reticulata* population.

Although *P. wingei* and *P. reticulata* share the same sex chromosome system (Darolti, et al. 2019; Morris, et al. 2018), we found more than twice as many sex-linked genes in *P. wingei* compared to *P. reticulata* (Table 1 and Table 2). Out of the total number of genes on the sex chromosome that were assigned a segregation type by SEX-DETector, only 92 (46.9%) are inferred as sex-linked in *P. reticulata*, compared to 249 (80.3%) in *P. wingei*. We also found 42 genes that are inferred as sex-linked in both species and that have orthologs in the outgroup species used in our phylogenetic analysis (Supplementary Table 4), representing approximately 46% and respectively 17% of the total number of sex-linked genes identified in *P. reticulata* and *P. wingei*.

Analysing the position of all genes along the sex chromosome revealed that sex-linked genes in *P. wingei* are spread throughout the entire length of the sex chromosome, while sex-linked genes in *P. reticulata* are predominantly found towards the distal arm of the chromosome (Figure 1, Supplementary Figure 2, Supplementary Figure 3). The differences we observe here are consistent with our previous findings based on male:female SNP density, that Stratum II in *P. wingei* appears to extend over a greater proportion of the sex chromosome compared to *P. reticulata* (Darolti, et al. 2019). Importantly, we also identified in both species the presence of a limited pseudoautosomal region at the distal end of the chromosome (26-27 Mb), where we recovered many genes with an autosomal inheritance but no sex-linked genes (Figure 1, Supplementary Figure 3), suggesting that recombination persists at a high rate in that area.

Compared to the previously defined pseudoautosomal region, we found a significant enrichment of *P. reticulata* sex-linked genes on the total predicted non-recombining region of the sex chromosome (Strata I and II together), as well as on each stratum independently (test done using the adjusted strata boundaries described below, Stratum I 15-21Mb, Stratum II 21-26Mb; $p < 0.02$ in all comparisons, Fisher's Exact test; Supplementary Table 5). The proportion of sex-linked genes, however, does not differ between the two strata in either of the two species ($p = 1$, odds ratio = 0.94, Fisher's exact test in *P. reticulata* and $p = 1$, odds

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

ratio = 1.03, Fisher's exact test in *P. wingei*). Approximately 36% of the genes on the *P. reticulata* pseudoautosomal region (0-15 Mb) were inferred to be sex-linked.

We used the 42 genes identified as sex-linked in both *P. reticulata* and *P. wingei*, together with orthologs in *P. formosa*, *Xiphophorus maculatus* and *Oryzias latipes*, in a phylogenetic analysis to investigate recombination suppression on the sex chromosomes. Both phylogenetic analyses reveal clustering of X- and Y-linked sequences by gametolog type instead of by species for four sex-linked genes, suggesting that these genes have stopped recombining before the two species diverged (Figure 2, Supplementary Figure 4). In *P. wingei*, these four sex-linked genes are all found in Stratum I (Darolti, et al. 2019), whereas in *P. reticulata* one of these genes is located in Stratum I, while the other three are found at the predicted boundary between the two strata (21-22Mb) (Wright, et al. 2017), suggesting that *P. reticulata* Stratum I is potentially wider than previously estimated based on sequence divergence alone. Considering this, for all our analyses here, we have adjusted the start point of the *P. reticulata* older stratum to 21Mb. However, in both species, we find that the majority of sex-linked genes cluster by species rather than by chromosome (Supplementary Figure 5, Supplementary Figure 6).

We estimated the rate of synonymous substitutions (dS_{XY}) between each pair of X- and Y-linked sequences. Mean dS_{XY} is >0 (mean dS_{XY} = 0.0059 in *P. reticulata*; mean dS_{XY} = 0.1433 in *P. wingei*; Supplementary Table 6, Supplementary Figure 7), and 95% CIs do not overlap with zero. However, we did not observe significant differences in dS_{XY} between the PAR, total non-recombining region, Stratum II and Stratum I in either species (Figure 3). We found no significant correlation between pairwise synonymous divergence of sex-linked genes and their position on the sex chromosome in either *P. reticulata* (r_s = 0.14, p = 0.279, Spearman's rank correlation) or *P. wingei* (r_s = 0.2, p = 0.095, Spearman's rank correlation). On average, genes identified as sex-linked in both species have a higher dS_{XY} in *P. wingei* than in *P. reticulata*, however this difference is only marginally significant (median dS_{XY} = 0.0045 in *P. reticulata*; median dS_{XY} = 0.0057 in *P. wingei*; p = 0.05, Wilcoxon signed rank test).

We also tested whether all identified Y-linked sequences have accumulated more deleterious mutations than X-linked sequences by estimating rates of nonsynonymous (d_N) and synonymous (d_S) substitutions and calculating average d_N/d_S for each gametolog branch.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Overall, Y-linked sequences in both species show a higher d_N/d_S compared to X-linked sequences, however this is not significant (Supplementary Table 6). These results are in line with our previous findings that neither of these species show Y degeneration or sex differences in transcription of genes on the sex chromosomes (Darolti, et al. 2019).

Compared to homologous X-linked genes, Y-linked genes are expected to gradually decrease in expression following recombination suppression (Chibalina and Filatov 2011). However, for either of the species, we found that average male Y/X expression ratio did not correlate with position on the sex chromosome (Figure 4; Adj. R^2 = -0.005, p = 0.376, linear regression slope = -0.010 for *P. reticulata,* Adj. R^2= -0.006, p = 0.774, linear regression slope = -0.002 for *P. wingei*) or with pairwise synonymous divergence (Figure 4; Adj. R^2 = -0.028, p = 0.518, linear regression slope = -0.002 for *P. reticulata,* Adj. R^2 = 0.004, p = 0.260, linear regression slope = -0.003 for *P. wingei*).

DISCUSSION

Complete recombination suppression over portions of the sex chromosomes is expected over time to favour the accumulation of deleterious mutations and lead to loss of gene activity and function on the sex-limited chromosome (Bachtrog 2013). In spite of these predictions, *P. reticulata* and *P. wingei* share a sex chromosome system that is largely non-recombining, yet which shows very low X-Y sequence divergence and no reduction of male gene activity and content (Darolti, et al. 2019; Wright, et al. 2017). These findings lead to questions about the mechanisms maintaining Y sequence integrity in these species. That these species also share a sex chromosome system with *P. picta*, which shows marked Y chromosome degeneration (Darolti, et al. 2019), suggests that the mechanism of recombination suppression is far less effective in *P. reticulata* and *P. wingei.*

Incomplete recombination suppression

In the complete absence of recombination, and with sufficient time for complete lineage sorting, we would expect phylogenetic analyses of sex-linked genes to show clustering of X- and Y-linked sequences by gametolog instead of by species. We have previously observed multiple concordant lines of evidence for two strata on the guppy sex chromosome. First, we see clear differences in male:female coverage for Stratum I and male:female SNP density for

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Stratum II in *P. reticulata* (Wright, et al. 2017). Second, we replicated these analyses in *P. wingei* and showed that these strata also exist in this species (Darolti, et al. 2019). Third, we see a conservation of male-specific *k*-mers between these species (Morris, et al. 2018), indicative of Y divergence in Stratum I. These strata were however not observed by Bergero, et al. (2019), most likely due to differences in mapping stringency (Wright, et al. 2019a).

Consistent with the evolutionary strata we have previously identified (Darolti, et al. 2019; Wright, et al. 2017), here, we observe signatures of regions with different time since recombination suppression. Specifically, we observe phylogenetic clustering by sex chromosome in four sex-linked genes (Figure 2, Supplementary Figure 4), suggesting that recombination suppression was finalized before the split of *P. wingei* and *P. reticulata*. Recombination suppression in the common ancestor is also consistent with the substantial amount of shared male-specific sequence that has been found in these species (Morris, et al. 2018). These genes are located either in Stratum I or at its predicted boundary with Stratum II (Supplementary Table 4), allowing us to define the sex chromosome region spanning 21-26 Mb as being the ancestral, more differentiated, non-recombining region. This is consistent with previous genetic mapping of the sex-determining locus to this region (Tripathi, et al. 2009), and the region where X-Y recombination has been previously undetected (Winge 1927, 1922; Yamamoto 1975).

Outside the older non-recombining region, in the previously identified Stratum II, we find that Y chromosomes in *P. reticulata* and *P. wingei* have a higher sequence similarity to their homologous X regions than to each other (Supplementary Figure 5, Supplementary Figure 6). This points to either incomplete lineage sorting or incomplete recombination suppression in this system. While recombination is largely curtailed in males (Bergero, et al. 2019), occasional recombination events prevent the rapid degeneration expected of fully non-recombining regions (Winge 1927, 1922; Yamamoto 1975). Infrequent X-Y recombination in males or in sex reversed females has been shown to be sufficient to maintain a high level of sequence similarity between gametologs in *Hyla* frogs for up to 7 million years (Stock, et al. 2011; Stock, et al. 2013), and has been documented in *P. reticulata* (Haskins, et al. 1961; Lisachov, et al. 2015) though it has been difficult to quantify the rate or the proportion of recombination nodules that result in strand invasion. That the guppy sex chromosome system is at least 20 million years old suggests that recombination suppression can remain permeable

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

on sex chromosomes far longer than previously recognized and can act as a long-term and persistent brake of Y chromosome degeneration.

It is also possible that we have underestimated the number of loci that have stopped recombining before *P. reticulata* and *P. wingei* split, as factors such as gene conversion or gene sequence length may cause a false signal of phylogenetic clustering by species. An alternative mechanism acting to prevent neutral degradation of the sex-limited chromosome is gene conversion between gametologs (Rosser, et al. 2009; Slattery, et al. 2000; Trombetta, et al. 2009; Wright, et al. 2014). Inter-chromosomal gene conversion events take place in meiosis when one allele is converted into its homolog in the process of double-strand DNA break repair (Chen, et al. 2007). Testing for the presence of gene conversion relies on identifying stretches of X-Y identical sequence delimited by variable sites (Sawyer 1999; Wright, et al. 2014). Unfortunately, here, the high similarity between the identified X- and Y-linked sequences in *P. reticulata* and *P. wingei* prevents such analyses.

We note that the lower phylogenetic bootstrap support values for the four single gene trees (Figure 2B-E) may be the result of incomplete lineage sorting, persistent gene conversion, and most importantly, from the short coding sequence of these loci. Bootstrap support for phylogenetic approaches such as this are often low, even for loci in significantly older strata (Wright, et al. 2014). Additionally, bootstrap support values are expected to be higher for majority consensus trees, which are constructed on alignments of many sets of genes, than for single gene trees. We observe that to be the case for both sex-linked genes that show clustering by gametolog (Figure 2) and for sex-linked genes that cluster by species (Supplementary Figure 5, Supplementary Figure 6).

We previously detected low levels of divergence in Stratum II (Darolti, et al. 2019; Wright, et al. 2017), and our analysis here also finds dS_{XY} >0 in this region (Figure 3). However, we previously detected significant accumulation of Y-specific SNPs (Darolti, et al. 2019; Wright, et al. 2017), suggesting that either recombinants are selected against, or that the rate of recombination is insufficient to fully counter the accumulation of Y mutations. It is worth mentioning that the permeability of recombination suppression found here can obscure additional differences between older and younger regions of recombination suppression.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

It is also possible that the degenerate *P. picta* Y chromosome represents the ancestral state, and the homomorphic system in *P. reticulata* and *P. wingei* represents a reversion to homomorphy, as has been recently documented in geckos (Rovatsos, et al. 2019). However, the presence of the complete X chromosome dosage compensation observed in *P. picta* (Darolti, et al. 2019) makes this unlikely. It has been argued that complete dosage compensation mechanisms will prevent sex chromosome turnover as the heterogametic sex would have two upregulated ancestral X chromosomes (Mank, et al. 2011; Van Doorn and Kirkpatrick 2010). Given the substantial gene content of the X chromosome (Kunstner, et al. 2016), this would lead to probable lethality due to over-expression of dosage sensitive genes in males.

Gradual expansion of recombination suppression

Our recent analyses of genomic coverage and polymorphism data in males and females indicate a larger non-recombining region in *P. wingei* compared to *P. reticulata* (Darolti, et al. 2019; Wright, et al. 2017). Indeed, here we find that the extent of sex-linkage is greater in *P. wingei* compared to *P. reticulata*. However, we also find sex-linked genes present in the previously defined *P. reticulata* pseudoautosomal region, suggesting that recombination suppression in this species may have expanded beyond the previously identified non-recombining region, but has not yet resulted in significant sequence divergence of the X and Y chromosomes. Alternatively, this may be due to partial linkage of the PAR when male recombination events occur somewhat distant from the PAR-sex chromosome boundary. We are also able to confirm the presence of a second small pseudoautosomal region at the distal end of the sex chromosomes (26-27Mb), region which has been previously suggested to be recombining (Bergero, et al. 2019; Darolti, et al. 2019; Lisachov, et al. 2015).

The results of our sex-linkage analysis suggest that, outside of Stratum I, either recombination suppression has occurred well after the split of *P. reticulata* and *P. wingei*, or recombination suppression is ancestral but has remained incomplete. Given that occasional recombination is still observed in this region (Winge 1927, 1922; Yamamoto 1975), the latter explanation is more likely, and this suggests that recombination suppression is imperfect, and likely not the result of large-scale inversions in this system. In the absence of an inversion, it is likely that the frequency or chromosomal location of these occasional recombination events vary across

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

different populations, consistent with the divergence differences we previously observed in natural populations of *P. reticulata* (Wright, et al. (2017) but see Bergero, et al. (2019) and Wright, et al. (2019a)). Furthermore, population-level differences in the rate of X-Y recombination could explain observed differences in the Y-linkage of guppy colour traits (Gordon, et al. 2012; Lindholm and Breden 2002).

Although we observe that $dS_{XY} > 0$, this is not due to significant increase in rates on the Y chromosome compared to the X (Supplementary Table 6), and does not differ across sex chromosome regions (Figure 3). Significant variation in pairwise synonymous substitution rates between strata are expected to be mainly seen in old sex chromosome systems where strata ages are very different. Previous studies on younger systems, such as those of *Mercurialis* (Veltsos, et al. 2019) and of *Silene* (Papadopulos, et al. 2015), show evidence of distinct evolutionary strata yet without significant differences in rates of gene sequence evolution between them. In completely non-recombining regions, the Y chromosome is expected to undergo a higher mutation rate than the X chromosome. However, incomplete recombination suppression, as observed here in the guppy system, will homogenize mutation rate between the X and the Y chromosomes such that we do not necessarily expect to see a consistently higher rate of mutation on the Y chromosome for all sex-linked genes. The results suggest a more gradual expansion of recombination suppression instead of a stepwise process resulting from inversions. It is worth noting though that discrete evolutionary strata and clear boundaries between the non-recombining region and the PAR may be masked by the stochastic differences between genes in X-Y divergence resulting from occasional recombination events (Chibalina and Filatov 2011).

Concluding remarks

Taken together, our results present a permeable picture of recombination suppression. The sex chromosomes in *P. reticulata* and *P. wingei* arose in the shared ancestral lineage with *P. picta* (Darolti, et al. 2019) roughly 20 million years ago (Meredith, et al. 2011) and have proceeded at markedly different rates. In *P. picta*, the highly degenerate Y chromosome is paired with complete X chromosome dosage compensation. However, in *P. reticulata* and *P. wingei*, which share substantial male-specific sequence (Morris, et al. 2018), there has been no perceptible loss of Y chromosome gene activity (Darolti, et al. 2019; Wright, et al. 2017).

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Our results indicate that occasional X-Y recombination acts to maintain Y chromosome integrity far longer than previously recognized (Stock, et al. 2011), and the degree to which recombination persists may explain the heterogeneity in the rate of Y chromosome degeneration observed across disparate sex chromosome systems (Bachtrog, et al. 2014).

ACKNOWLEDGEMENTS

This work was supported by the European Research Council (grant agreement 680951 to J.E.M.), the Biotechnology and Biological Sciences Research Council (PhD to I.D., grant number BB/M009513/1) and JEM gratefully acknowledges support from a Canada 150 Research Chair. We acknowledge the use of the UCL Legion High-Performance Computing Facility (Legion@UCL) in the completion of this work. We thank P. Almeida, A. Corral-Lopez, L. Fong, B. Furman, Y. Lin, D. Metzger, J. Morris, L. Rowe, B. Sandkam, J. Shu and W. van der Bijl for helpful comments and suggestions.

AUTHOR CONTRIBUTIONS

J.E.M. and I.D. designed the research; I.D. established the poeciliid families and performed the RNA extractions; I.D. performed the research; I.D., A.E.W. and J.E.M. analyzed the data; I.D., A.E.W. and J.E.M. wrote the manuscript.

DATA ACCESSIBILITY

RNA-seq reads have been deposited at the NCBI Sequencing Read Archive (BioProject ID PRJNA591249).

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

REFERENCES

Almeida P, et al. 2019. Single-molecule genome assembly of the Basket Willow, Salix viminalis, reveals earliest stages of sex chromosome expansion. *bioRxiv*: 589804.

Altschul SF, Gish W, Miller W, Myers E, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3): 403-410.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 14(2): 113-124.

Bachtrog D, et al. 2011. Are all sex chromosomes created equal? *Trends Genet*. 27(9): 350-357.

Bachtrog D, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol*. 12(7): e1001899.

Bergero R, Gardner J, Bader B, Yong L, Charlesworth D. 2019. Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proc Natl Acad Sci USA*. 116(14): 6924-6931.

Bergero R, Qiu S, Forrest A, Borthwick H, Charlesworth D. 2013. Expansion of the pseudo-autosomal region and ongoing recombination suppression in the Silene latifolia sex chromosomes. *Genetics*. 194(3): 673-686.

Bloch NI, et al. 2018. Early neurogenomic response associated with variation in guppy female mate preference. *Nat Ecol Evol*. 2(11): 1772-1781.

Bull JJ. 1983. Evolution of sex determining mechanisms: The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Campos J, Qiu S, Guirao-Rico S, Bergero R, Charlesworth D. 2017. Recombination changes at the boundaries of fully and partially sex-linked regions between closely related Silene species pairs. *Heredity*. 118(4): 395-403.

Charlesworth B, Charlesworth De. 2000. The degenration of Y chromosomes. *Philos Trans R Soc B*. 355(1403): 1563-1572.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity*. 95(2): 118-128.

Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet*. 8(10): 762-775.

Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr Biol*. 21(17): 1475-1479.

Cortez D, et al. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 508(7497): 488-493.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

564 Darolti I, et al. 2019. Extreme heterogeneity in sex chromosome differentiation and
565 dosage compensation in livebearers. *Proc Natl Acad Sci USA*. 116(38): 19031-19036.

566 Flicek P, et al. 2014. Ensembl 2014. *Nucleic Acids Res.* 42(D1): 749-755.

567 Gordon SP, Lopez-Sepulcre A, Reznick DN. 2012. Predation-associated differences
568 in sex linkage of wild guppy coloration. *Evolution*. 66(3): 912-918.

569 Grabherr MG, et al. 2011. Trinity: reconstructing a full-length transcriptome without a
570 genome from RNA-Seq data. *Nat Biotechnol*. 29(7): 644-652.

571 Handley L-JL, Ceplitis H, Ellegren H. 2004. Evolutionary strata on the chicken Z
572 chromosome: implications for sex chromosome evolution. *Genetics*. 167(1): 367-376.

573 Harris RS. 2007. Improved pairwise alignment of genomic DNA. PhD thesis,
574 Pennsylvania State University, University Park, PA.

575 Harrison PW, Jordan GE, Montgomery SH. 2014. SWAMP: sliding window alignment
576 masker for PAML. *Evol Bioinform*. 10: EBO. S18193.

577 Harrison PW, et al. 2015. Sexual selection drives evolution and rapid turnover of male
578 gene expression. *Proc Natl Acad Sci USA*. 112(14): 4393-4398.

579 Haskins CP, Haskins EF, McLaughlin J, Hewitt R. 1961. Polymorphism and population
580 structure in Lebistes reticulatus, an ecological study. *Vertebrate speciation*. 320: 395.

581 Hough J, Hollister JD, Wang W, Barrett SC, Wright SI. 2014. Genetic degeneration of
582 old and young Y chromosomes in the flowering plant Rumex hastatulus. *Proc Natl
Acad Sci USA*. 111(21): 7713-7718.

584 Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res*.
585 9(9): 868-877.

586 Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software
587 platform for the organization and analysis of sequence data. *Bioinformatics*. 28(12):
588 1647-1649.

589 Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron:
590 duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl
Acad Sci USA*. 100(20): 11484-11489.

592 Kotrschal A, et al. 2013. Artificial selection on relative brain size in the guppy reveals
593 costs and benefits of evolving a larger brain. *Curr Biol*. 23(2): 168-171.

594 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary
595 genetics analysis across computing platforms. *Mol Biol Evol*. 35(6): 1547-1549.

596 Kunstner A, et al. 2016. The genome of the Trinidadian guppy, Poecilia reticulata, and
597 variation in the Guanapo population. *PLoS ONE*. 11(12): e0169087.

598 Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome.
Science. 286(5441): 964-967.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3): R25.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 12(1): 323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14): 1754-1760.

Li H, et al. 2009. The sequence alignment / map format and SAMtools. *Bioinformatics*. 25(16): 2078-2079.

Lindholm A, Breden F. 2002. Sex chromosomes and sexual selection in poeciliid fishes. *Am Nat*. 160(S6): 214-224.

Lisachov AP, Zadesenets KS, Rubtsov NB, Borodin PM. 2015. Sex chromosome synapsis and recombination in male guppies. *Zebrafish*. 12(2): 174-180.

Lohse M, et al. 2012. RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40(W1): 622-627.

Löytynoja A, Goldman N. 2008. Phylogeny-aware gap replacement prevents errors in sequence alignment and evolutionary analysis. *Science*. 320(5883): 1632-1635.

Mank JE, Hosken DJ, Wedell N. 2011. Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. *Evolution*. 65(8): 2133-2144.

Martin H, et al. 2019. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. *Genome Biol Evol*. 11(2): 350-361.

Meredith RW, Pires MN, Reznick DN, Springer MS. 2011. Molecular phylogenetic relationships and the coevolution of placentotrophy and superfetation in Poecilia (Poeciliidae: Cyprinodontiformes). *Mol Phylogenet Evol*. 59(1): 148-157.

Morris J, Darolti I, Bloch NI, Wright AE, Mank JE. 2018. Shared and Species-Specific Patterns of Nascent Y Chromosome Evolution in Two Guppy Species. *Genes (Basel)*. 9(5): 238.

Muyle A, et al. 2016. SEX-DETector: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol Evol*. 8(8): 2530-2543.

Muyle A, Shearn R, Marais GAB. 2017. The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol Evol*. 9(3): 627-645.

Muyle A, et al. 2012. Rapid de novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. *PLoS Biol.* 10(4): e1001308.

Muyle A, et al. 2018. Genomic imprinting mediates dosage compensation in a young plant XY system. *Nature plants*. 4(9): 677-680.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

636 Natri HM, Shikano T, Merila J. 2013. Progressive recombination suppression and
637 differentiation in recently evolved neo-sex chromosomes. *Mol Biol Evol*. 30(5): 1131-
638 1144.

639 Nicolas M, et al. 2004. A gradual process of recombination restriction in the
640 evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* 3(1): e4.

641 Papadopulos AS, Chester M, Ridout K, Filatov DA. 2015. Rapid Y degeneration and
642 dosage compensation in plant sex chromosomes. *Proceedings of the National*
643 *Academy of Sciences*. 112(42): 13021-13026.

644 R Core Team. 2015. R: A language and environment for statistical computing. *Vienna,*
645 *Austria*. https://www.R-project.org/.

646 Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback
647 genome--patterns and consequences. *Mol Ecol*. 22(11): 3014-3027.

648 Ross MT, et al. 2005. The DNA sequence of the human X chromosome. *Nature*.
649 434(7031): 325-337.

650 Rosser ZH, Balaresque P, Jobling MA. 2009. Gene conversion between the X
651 chromosome and the male-specific region of the Y chromosome at a translocation
652 hotspot. *Am J Hum Genet*. 85(1): 130-134.

653 Rovatsos M, Farkačová K, Altmanová M, Johnson Pokorná M, Kratochvíl L. 2019. The
654 rise and fall of differentiated sex chromosomes in geckos. *Mol Ecol*. 28: 3042-3052.

655 Sawyer SA. 1999. GENECONV: a computer package for the statistical detection of
656 gene conversion. Distributed by the author www.math.wustl.edu/sawyer/geneconv/.

657 Slattery JP, Murphy WJ, O'Brien SJ. 2000. Patterns of diversity among SINE elements
658 isolated from three Y-chromosome genes in carnivores. *Mol Biol Evol*. 17(5): 825-829.

659 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
660 analysis of large phylogenies. *Bioinformatics*. 30(9): 1312-1313.

661 Stock M, et al. 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biol*.
662 9: e1001062.

663 Stock M, et al. 2013. Low rates of X-Y recombination, not turnovers, account for
664 homomorphic sex chromosomes in several diploid species of Palearctic green toads
665 (Bufo viridis subgroup). *J Evol Biol*. 26(3): 674-682.

666 Tripathi N, et al. 2009. Genetic linkage map of the guppy, Poecilia reticulata, and
667 quantitative trait loci analysis of male size and colour variation. *Proc Biol Sci*.
668 276(1665): 2195-2208.

669 Trombetta B, Cruciani F, Underhill PA, Sellitto D, Scozzari R. 2009. Footprints of X-
670 to-Y gene conversion in recent human evolution. *Mol Biol Evol*. 27(3): 714-725.

671 Van Doorn GS, Kirkpatrick M. 2010. Transitions between male and female
672 heterogamety caused by sex-antagonistic selection. *Genetics*. 186(2): 629-645.

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

673 Veltsos P, et al. 2019. Early sex-chromosome evolution in the diploid dioecious plant
674 Mercurialis annua. *Genetics*. 212(3): 815-835.

675 Winge Ö. 1927. The location of eighteen genes in Lebistes reticulatus. *J Genet*. 18(1):
676 1-43.

677 Winge Ö. 1922. One-sided masculine and sex-linked inheritance in Lebistes
678 reticulatus. *J Genet*. 12: 145-162.

679 Wright AE, et al. 2017. Convergent recombination suppression suggests role of sexual
680 selection in guppy sex chromosome formation. *Nat Commun*. 8: 14251.

681 Wright AE, et al. 2019a. On the power to detect rare recombination events. *Proc Natl*
682 *Acad Sci USA*. 116(26): 12607-12608.

683 Wright AE, Harrison PW, Montgomery SH, Pointer MA, Mank JE. 2014. Independent
684 stratum formation on the avian sex chromosomes reveals inter-chromosomal gene
685 conversion and predominance of purifying selection on the W chromosome. *Evolution*.
686 68(11): 3281-3295.

687 Wright AE, Moghadam HK, Mank JE. 2012. Trade-off between selection for dosage
688 compensation and masculinization on the avian Z chromosome. *Genetics*. 192(4):
689 1433-1445.

690 Wright AE, Rogers TF, Fumagalli M, Cooney CR, Mank JE. 2019b. Phenotypic sexual
691 dimorphism is associated with genomic signatures of resolved sexual conflict. *Mol*
692 *Ecol*. 28(11): 2860-2871.

693 Yamamoto T. 1975. The medaka, Oryzias latipes, and the guppy, Lebistes reticularis.
694 In. Handbook of genetics: Springer. p. 133-149.

695 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*.
696 24(8): 1586-1591.

697 Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution
698 rates under realistic evolutionary models. *Mol Biol Evol*. 17(1): 32-43.

699

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

TABLES

Table 1. Number of inferred *P. reticulata* sex-linked genes. These numbers consist of both X/Y and X0 genes.

Sex-linked genes	All families	Family 1	Family 2	Family 3
Total number	111	60	41	28
On the sex chromosomes	92 (83%)	54 (90%)	33 (81%)	21 (75%)
On the autosomes	13 (12%)	5 (8%)	3 (7%)	5 (18%)
On unplaced scaffolds	6 (5%)	1 (2%)	5 (12%)	2 (7%)

Table 2. Number of inferred *P. wingei* sex-linked genes. These numbers consist of X/Y gene pairs only as there were no identified X0 genes in *P. wingei*.

Sex-linked genes	All families	Family 1	Family 2
Total number	272	236	179
On the sex chromosomes	249 (92%)	219 (93%)	163 (91%)
On the autosomes	12 (4%)	7 (3%)	7 (4%)
On unplaced scaffolds	11 (4%)	10 (4%)	9 (5%)

Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

FIGURES

Figure 1. Density of sex-linked genes (top panels) and all genes (bottom panels) across the sex chromosomes of *P. reticulata* (A) and *P. wingei* (B). The shaded purple regions indicate the previously identified non-recombining regions (Darolti, et al. 2019; Wright, et al. 2017). Stratum I is shown in dark purple, where X-Y divergence is the greatest, and Stratum II is shown in light purple.

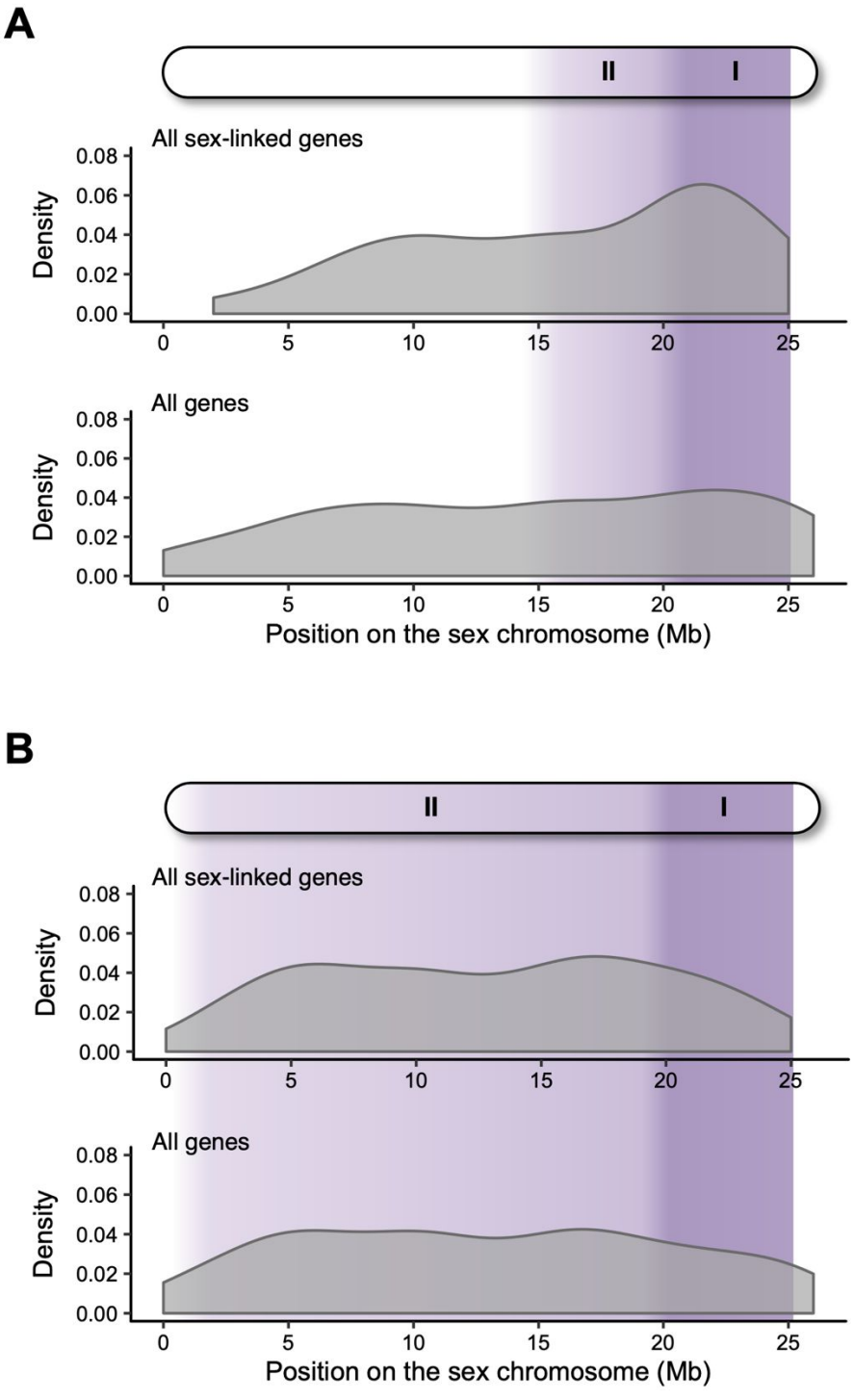
Figure 2. Phylogenetic gene trees for *P. reticulata* and *P. wingei* X- and Y-linked sequences. Phylogenetic trees for the four sex-linked genes in which the X (red) and Y (blue) sequences cluster by gametolog instead of by species. (A) Consensus tree based on alignments of all four sex-linked genes. Numbers at each node represent bootstrap values based on 100 permutations. Branches with the interrupted lines have been shortened to improve clarity. (B) *alad* (C) *dnajc25* (D) *LOC103473940* (E) *LOC103474035*.

Figure 3. Pairwise synonymous divergence (dS_{XY}) of *P. reticulata* (A) and *P. wingei* (B) sex-linked genes across the sex chromosomes. The shaded purple regions indicate the non-recombining regions. Stratum I is shown in dark purple, where X-Y divergence is the greatest, and Stratum II is shown in light purple. Lines show linear regressions fitted to the data using the *lm* function in R, and grey shaded areas represent confidence intervals for the slope of the regression lines.

Figure 4. Average male Y/X expression ratio plotted against sex-linked gene position on the sex chromosome (A) and pairwise synonymous divergence (dS_{XY}) (B). *P. reticulata* data is represented in orange, while *P. wingei* data is in blue. Lines show linear regressions fitted to the data using the *lm* function in R, and orange and blue shaded areas represent confidence intervals for the slope of the regression lines.

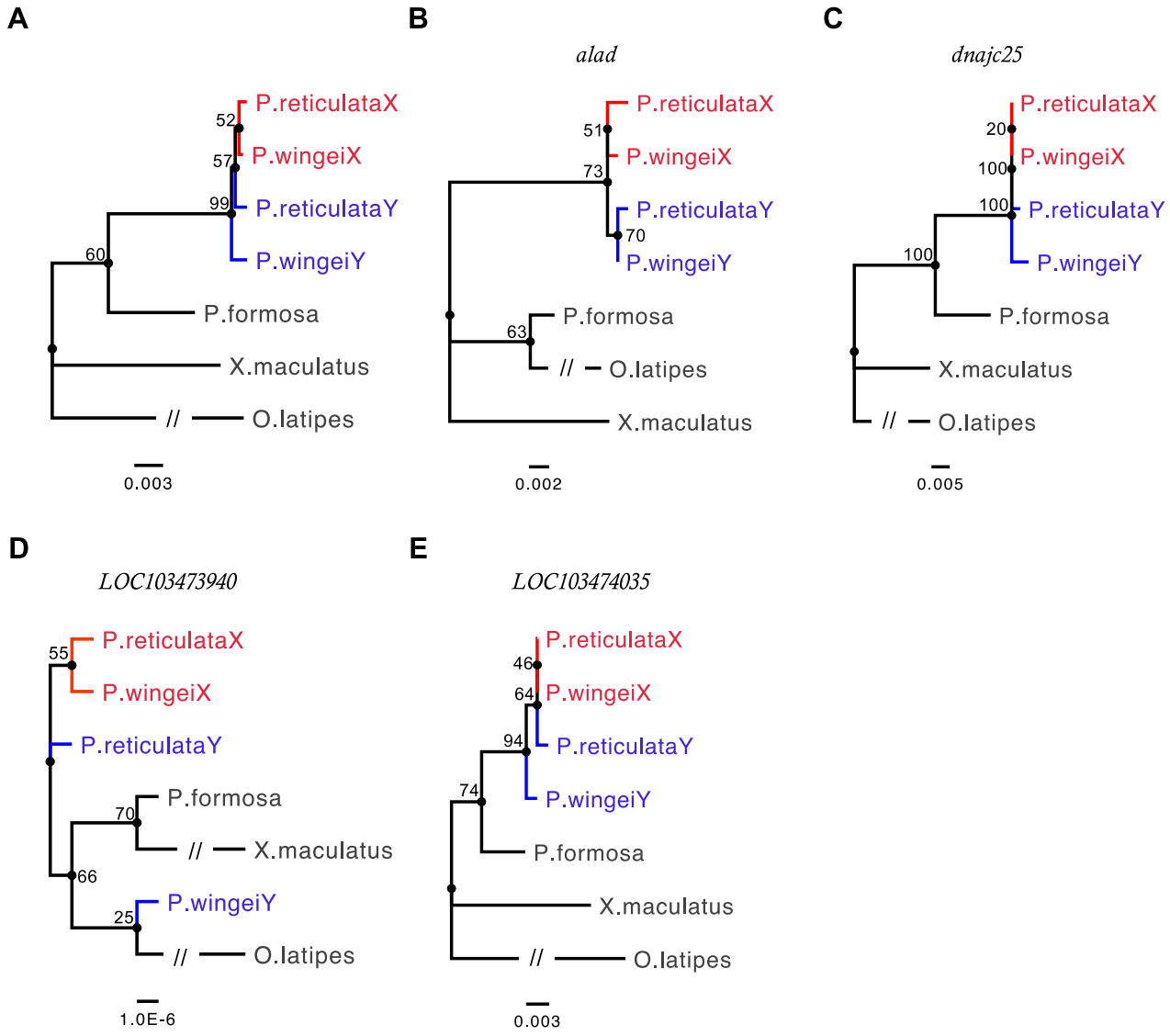
Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Figure 1



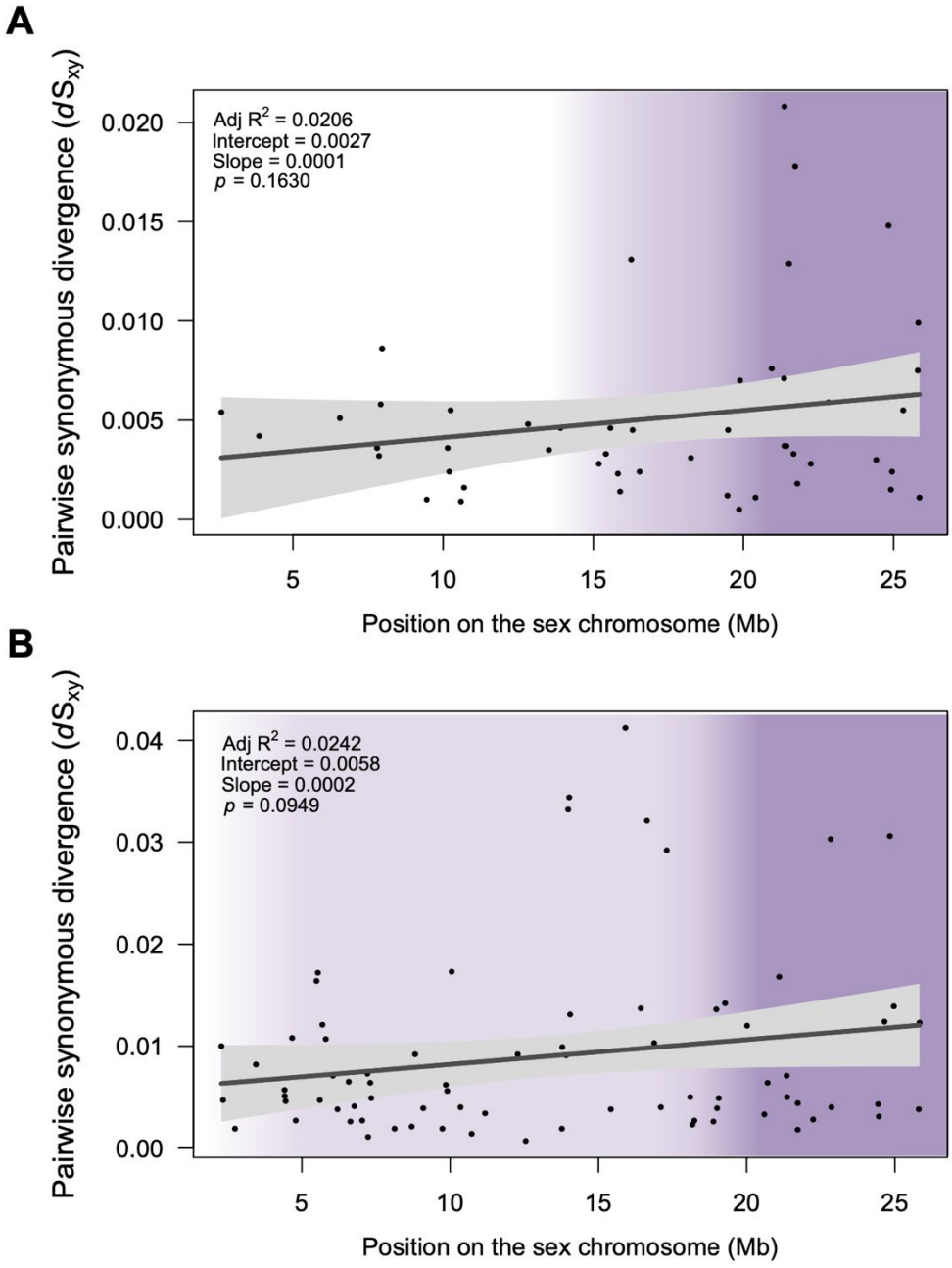
Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Figure 2



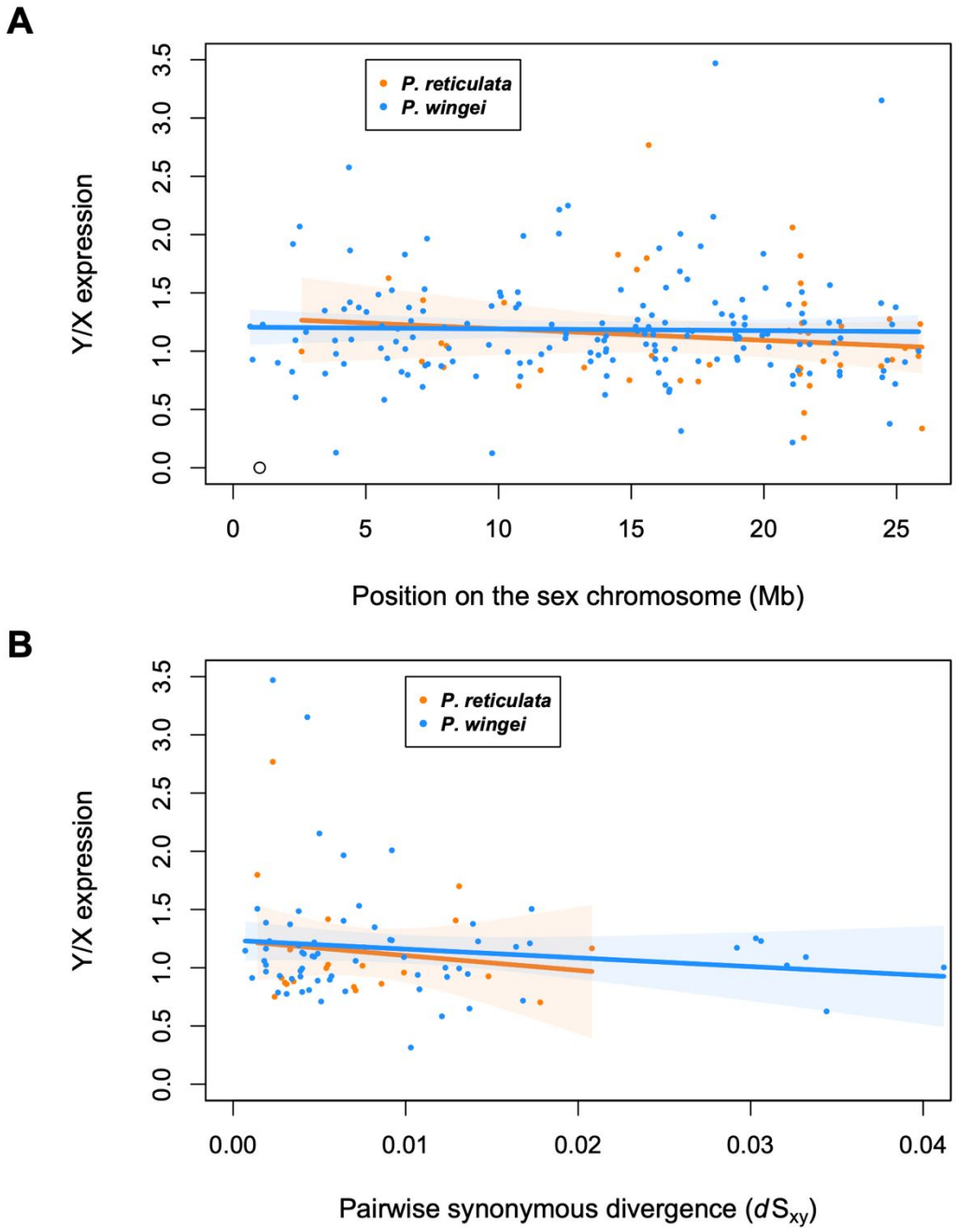
Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Figure 3



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020

Figure 4



Downloaded from https://academic.oup.com/gbe/advance-article-abstract/doi/10.1093/gbe/evaa099/5840475 by guest on 31 May 2020