

Positive Selection Drives *cis*-regulatory Evolution Across the Threespine Stickleback Y Chromosome

Daniel E. Shaw ¹, Alice Shanfelter Naftaly ¹ and Michael A. White ^{1,*}

¹Department of Genetics, University of Georgia, Athens, GA 30602, USA

*Corresponding author: E-mail: whitem@uga.edu.

Associate editor: Patricia Wittkopp

Abstract

Allele-specific gene expression evolves rapidly on heteromorphic sex chromosomes. Over time, the accumulation of mutations on the Y chromosome leads to widespread loss of gametolog expression, relative to the X chromosome. It remains unclear if expression evolution on degrading Y chromosomes is primarily driven by mutations that accumulate through processes of selective interference, or if positive selection can also favor the down-regulation of coding regions on the Y chromosome that contain deleterious mutations. Identifying the relative rates of *cis*-regulatory sequence evolution across Y chromosomes has been challenging due to the limited number of reference assemblies. The threespine stickleback (*Gasterosteus aculeatus*) Y chromosome is an excellent model to identify how regulatory mutations accumulate on Y chromosomes due to its intermediate state of divergence from the X chromosome. A large number of Y-linked gametologs still exist across 3 differently aged evolutionary strata to test these hypotheses. We found that putative enhancer regions on the Y chromosome exhibited elevated substitution rates and decreased polymorphism when compared to nonfunctional sites, like intergenic regions and synonymous sites. This suggests that many *cis*-regulatory regions are under positive selection on the Y chromosome. This divergence was correlated with X-biased gametolog expression, indicating the loss of expression from the Y chromosome may be favored by selection. Our findings provide evidence that Y-linked *cis*-regulatory regions exhibit signs of positive selection quickly after the suppression of recombination and allow comparisons with recent theoretical models that suggest the rapid divergence of regulatory regions may be favored to mask deleterious mutations on the Y chromosome.

Key words: positive selection, *cis*-regulatory elements, sex chromosomes, stickleback fish.

Introduction

The evolution of heteromorphic sex chromosomes has occurred many times among species (Bachtrog 2013; Bachtrog et al. 2014). Heteromorphic sex chromosomes evolve once recombination is suppressed between the X and Y (or Z and W) (Muller 1918; Charlesworth 1978; Charlesworth and Charlesworth 2000). After recombination is suppressed, the Y chromosome rapidly accumulates mutations, leading to sequence degeneration (Charlesworth and Charlesworth 2000). Empirical evidence of this process has focused on the sequence evolution of coding regions on sex chromosomes. Broad comparative work has revealed that coding sequence evolution can follow one of several different evolutionary trajectories. Although many ancestral Y-linked genes are lost because of sequence degeneration through the accumulation of deleterious mutations, growing evidence indicates not all genes are lost. Some genes appear to be dosage sensitive and are under strong purifying selection to be retained on the Y chromosome (Bellott et al. 2014, 2017; White et al. 2015; Peichel et al. 2020). Novel genes with sex-specific functions can also accumulate on sex-limited

chromosomes via translocation and subsequent gene duplication (Soh et al. 2014; Mahajan and Bachtrog 2017; Ellison and Bachtrog 2019; Hughes et al. 2020; Peichel et al. 2020; Chang et al. 2022).

The accumulation of mutations is not restricted to coding sequences and can also occur in *cis*-regulatory regions, leading to changes in expression from the sex-limited chromosome. Gametolog expression (ancestral genes still shared between the X and Y chromosomes) has been shown to evolve rapidly on degenerating sex chromosomes, often resulting in the loss of expression from the sex-limited chromosome (Y and W) (Meisel et al. 2012a; Muyle et al. 2012, 2018; Ayers et al. 2013; Singh et al. 2014; White et al. 2015; Beaudry et al. 2017; Rodríguez Lorenzo et al. 2018; Martin et al. 2019; Veltsos et al. 2019; Wei and Bachtrog 2019; Shaw and White 2022). Although deleterious regulatory mutations may accumulate through selective interference (Charlesworth and Charlesworth 2000; Bachtrog 2008), selection may also favor mutations within *cis*-regulatory regions to downregulate coding regions with deleterious mutations (Orr and Kim 1998; Bachtrog 2006). Recent theory supports the role of positive selection driving the rapid accumulation of mutations to

Received: August 11, 2023. Revised: December 31, 2023. Accepted: January 24, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

downregulate the Y-linked allele and upregulate the X-linked allele to maintain ancestral dosage balance (Lenormand et al. 2020; Lenormand and Roze 2022).

One challenge to characterizing the molecular evolution of regulatory regions on sex chromosomes is that complete Y chromosome assemblies are only available for a limited number of taxa. Reference assemblies are needed to identify the substitutions that are accumulating within noncoding regions. Additionally, many of the available Y chromosome assemblies are from model organisms that have sex chromosomes that are highly degenerated (Hughes et al. 2010, 2020; Bellott et al. 2014; Soh et al. 2014; Tomaszewicz et al. 2016; Mahajan et al. 2018). These species only have a few remaining ancestral gametologs on the Y chromosome, thus limiting the number of genes available to study how *cis*-regulatory evolution potentially leads to expression differences between the X and Y. Species with recently derived sex chromosomes that still harbor many ancestral Y-linked gametologs at varied stages of degeneration are needed to understand how substitutions within *cis*-regulatory regions lead to the evolution of expression differences between the X and Y.

In addition to having a chromosome-scale assembly of the Y chromosome, an additional challenge is annotating regulatory regions (reviewed in Shaw and White 2022). Recent approaches have focused on identifying regions of the genome with accessible chromatin that may contain transcription factor binding sites to regulate gene expression (Ricci et al. 2019). Chromatin accessibility profiling techniques, like assay for transposase accessible chromatin sequencing (ATAC-seq) (Buenrostro et al. 2013), utilize short-read sequencing to profile accessible chromatin regions (ACRs) at a fine scale across the genome. Across autosomes, ACRs show signatures of purifying selection relative to other intergenic regions, consistent with these regions containing important functional elements for gene regulation (Connelly et al. 2014; Lu et al. 2019; Horvath et al. 2021). Profiling ACRs on sex chromosomes would therefore provide a means to study the molecular evolution of *cis*-regulatory regions in the context of Y degeneration.

The threespine stickleback fish (*Gasterosteus aculeatus*) is an excellent model to study the evolution of *cis*-regulatory regions on sex chromosomes. Threespine stickleback fish have a high-quality reference assembly of the Y chromosome and this chromosome is more recently derived compared to many other species with available chromosome-scale assemblies (Hughes et al. 2010, 2012, 2020; Bellott et al. 2014, 2017; Peichel et al. 2020). The threespine stickleback sex chromosomes contain multiple evolutionary strata at different stages of degeneration (Roesti et al. 2013; White et al. 2015; Peichel et al. 2020), allowing for comparisons of sequence evolution at different temporal scales. Strata form if recombination is suppressed between sex chromosomes in multiple steps (Bachtrog 2013). Each subsequent phase of recombination suppression is followed by the accumulation of substitutions on the Y chromosome, leading to increased sequence divergence between the X and Y. The age of individual strata can therefore be distinguished by

varied levels of synonymous divergence. In threespine stickleback fish, the oldest stratum is estimated to have formed around 22 million years ago. Two younger strata formed between 4 and 6 million years ago (Peichel et al. 2020; Sardell et al. 2021). Stratum 1 has undergone the most extensive degeneration, with only 18% of ancestral coding regions remaining on the Y chromosome (Peichel et al. 2020). Many of the coding regions of these gametologs that remain in stratum 1 on the Y chromosome show signatures of purifying selection and are enriched for dosage-sensitive functions (White et al. 2015; Peichel et al. 2020). Interestingly, chromosome-wide dosage compensation has not evolved to counter the loss of expression for genes where the coding sequence has degenerated on the Y chromosome (White et al. 2015). The lack of chromosome-wide dosage compensation enables a gene-by-gene comparison of *cis*-regulatory changes to understand how expression of X-linked alleles evolves in response to degenerating coding regions on the Y.

Here, we leveraged ATAC-seq from 2 different tissues to identify *cis*-regulatory regions on the sex chromosomes and an autosome. We found sex-linked *cis*-regulatory regions exhibited signatures of positive selection, relative to nonfunctional, intergenic control regions. Our findings complement existing sex chromosome evolution theory that suggests the accumulation of mutations within *cis*-regulatory regions may be beneficial to silence expression from a degenerating Y chromosome.

Results

Nucleotide Substitutions are Accumulating at High Rates Within Accessible Chromatin Regions

The completion of the threespine stickleback Y chromosome assembly (Peichel et al. 2020) allowed us to thoroughly examine regulatory evolution among sex-linked gametologs shared between the sex chromosomes. ACRs putatively contain *cis*-regulatory elements that serve as domains for transcription factor binding. High sequence divergence within ACRs, therefore, represents regions of the genome that are likely experiencing *cis*-regulatory evolution. We first sought to compare the divergence of ACRs between the X and Y chromosomes to previously characterized synonymous divergence throughout coding regions. Crossing over was suppressed across the threespine stickleback sex chromosomes in at least 3 separate events, forming strata with distinct levels of divergence (stratum 1: oldest; stratum 2: intermediate; stratum 3: youngest) (Peichel et al. 2020). To survey ACRs, we utilized ATAC-seq from liver tissue (Naftaly et al. 2021) and testis tissue. We defined a set of noncoding ACRs that were likely present in the ancestor of the X and Y chromosomes, by identifying X-linked ACRs within intergenic sequence that still had homologous sequence on the Y chromosome and homologous sequence in the ancestral autosome (homologous sequence in ninespine stickleback chromosome 19). To calculate divergence, we compared these regions with the homologous sequence on the Y chromosome, which may no longer be associated with accessible chromatin. In the liver, we found a total of 1,279

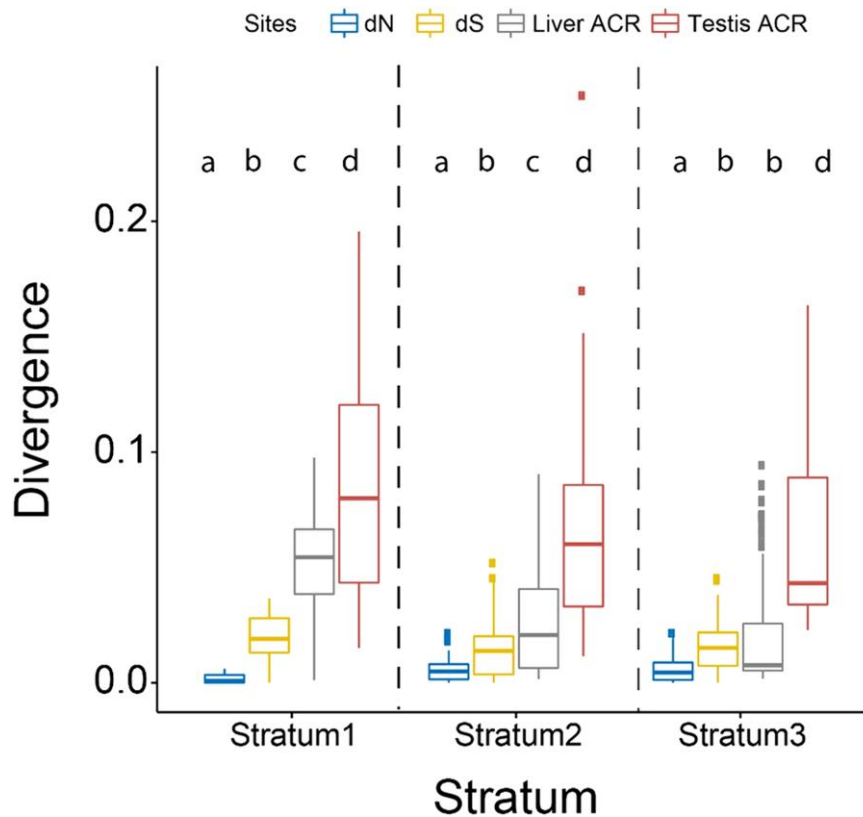


Fig. 1. XY divergence within ACRs and coding regions. ACRs show elevated divergence between the X and Y, relative to nonsynonymous and synonymous substitutions in coding regions. Protein-coding divergence was estimated between the X and Y chromosomes at all shared gametologs. ACRs were identified on the X chromosome and mapped to the Y chromosome. Substitutions within ACRs were summed and divided by the total length of the ACR to calculate divergence. Stratum 1: 12 gametologs, 442 liver ACRs, 30 testis ACRs; Stratum 2: 71 gametologs, 228 liver ACRs, 37 testis ACRs; Stratum three: 46 gametologs, 199 liver ACRs 51 testis ACRs. Rates of sequence divergence for all 4 categories were compared using a Kruskal–Wallis test for each stratum ($P < 0.001$ for all strata). A post hoc Dunn's test was used to identify which groups were significantly different from one another. P -values were adjusted for multiple comparisons with a Bonferroni correction. The letters above each plot indicate the significant differences between sites within each stratum ($P < 0.001$). Box plots that share a letter assignment are not significantly different from one another ($P > 0.001$).

X-linked ACRs. We were able to align 948 (74.1%) of these ACRs to homologous regions on the Y chromosome and the ancestral ninespine stickleback (*Pungitius pungitius*) autosome 19. 869 liver ACRs were within 50 kb of a coding region shared by all three chromosomes. In the testis, we identified 896 X-linked ACRs. Compared to the liver, we found far fewer orthologous regions on the Y chromosome and autosome 19 in the ninespine stickleback (118 total aligned; 13.2%), suggesting *cis*-regulatory regions within testis ACRs may be more rapidly evolving. All 118 of the aligned testis ACRs were within 50 kb of a coding region shared by all 3 chromosomes. We used these aligned sets of ACRs that were close to conserved coding regions for all subsequent analyses.

We estimated divergence between the X and Y chromosomes among ACRs in each tissue and compared this to synonymous and nonsynonymous substitutions within coding regions across the sex chromosomes. Coding regions across all 3 strata on the threespine stickleback Y chromosome are accumulating nonsynonymous substitutions at a higher rate than autosomes due to reduced efficacy of selection (White et al. 2015; Peichel et al. 2020). We first compared

the divergence among testis and liver ACRs to nonsynonymous substitution rates within coding regions to determine if mutations were accumulating at a similar rate. We found liver and testis ACRs have much higher divergence compared to nonsynonymous mutations among all 3 strata (Fig. 1) ($P < 0.001$; Kruskal–Wallis and Dunn's Test). We also compared ACR divergence with synonymous substitutions, which is used as measure of neutral evolution. Testis ACRs had divergence higher than neutral, synonymous substitutions in all 3 strata (Fig. 1) ($P < 0.001$; Kruskal–Wallis and Dunn's Test). Liver ACRs exhibited divergence that was higher than synonymous substitutions in the 2 oldest strata (1 and 2) ($P < 0.001$; Kruskal–Wallis and Dunn's test) (Fig. 1), indicating these regions are also evolving rapidly. In the youngest stratum (3), the liver ACRs were not significantly different than synonymous substitutions ($P = 0.413$, Kruskal–Wallis and Dunn's test). Additionally, ACR divergence was distinct between all 3 evolutionary strata ($P < 0.001$; Kruskal–Wallis and Dunn's test) and corresponded with the extent of evolutionary divergence of each stratum. ACR divergence higher than both nonsynonymous and synonymous substitutions strongly suggests *cis*-regulatory

regions are evolving rapidly across all 3 strata. However, this analysis alone cannot rule out the possibility that purifying selection may still be operating within coding regions on the Y chromosome, reducing nonsynonymous substitutions. Synonymous substitutions may also be under weak purifying selection (i.e. codon bias) (Bachtrog 2005; Lawrie et al. 2013). In this scenario, we may observe higher divergence rates within ACRs relative to coding regions.

We, therefore, sought to compare ACR divergence with nonfunctional regions throughout the sex chromosomes as well as ACR divergence throughout a representative autosome. We used the orthologous sequence from the ninespine stickleback fish to infer substitution rates throughout the X and Y chromosomes separately as well as throughout autosome 18. We defined nonfunctional control regions as intergenic sequence of the chromosomes that did not overlap known gene annotations, repetitive elements, or our previously defined ACRs (see Materials and Methods). Although our ability to define nonfunctional regions is limited to the currently available annotations, this represents our best approximation of regions of the genome that are not under selection. If ACRs are evolving faster than expected under neutrality, we would predict the average ACR sequence divergence would exceed the rate of divergence observed across permutations of nonfunctional control regions (Fig. 2a). However, if ACR divergence is similar to the control regions (Fig. 2a), this would suggest substitutions are accumulating in ACRs at the same rate as the remainder of the Y chromosome, which is subject to sequence degeneration across the noncrossover region (Charlesworth and Charlesworth 2000).

ACRs functional in the liver exhibited higher substitution rates on the Y chromosome than on the X chromosome. Y-linked ACRs had significantly higher substitution rates compared to the nonfunctional control regions for the 2 oldest evolutionary strata (Fig. 2b; $P < 0.001$; stratum 1 and 2; 10,000 permutations). In contrast to the Y-linked ACRs, we found the X-linked liver ACR sequence divergence was significantly lower than nonfunctional control regions (Fig. 2b; Stratum 1: $P = 0.016$; Stratum 2 and 3: $P < 0.001$; 10,000 permutations). This suggests that ACRs may be functionally constrained more on the X chromosome, compared to Y-linked ACRs. We observed a similar pattern on the Y chromosome for ACRs functional in testis tissue. Y-linked ACRs had significantly higher substitution rates in all 3 evolutionary strata, compared to nonfunctional control regions (Fig. 2b; $P < 0.001$, 10,000 permutations). However, we observed X-linked ACRs from testis tissue, also exhibited elevated nucleotide substitutions in the oldest 2 evolutionary strata (Fig. 2b; $P < 0.001$; 10,000 permutations), while the youngest stratum was significantly lower than intergenic regions (Fig. 2b; $P < 0.001$; 10,000 permutations). This indicates some *cis*-regulatory elements exhibited accelerated substitution rates on the X chromosome and this occurs in a tissue-specific fashion.

Autosomal ACRs have been shown to be under purifying selection in other species (Horvath et al. 2021). We therefore

tested whether threespine stickleback autosomal ACRs also exhibited purifying selection, similar to the pattern we observed on the X chromosome in liver. Consistent with patterns of purifying selection, we found autosomal ACRs exhibited lower sequence divergence compared to nonfunctional control regions in both liver (Fig. 2b; $P = 0.002$; 10,000 permutations) and testis tissue (Fig. 2b; $P < 0.001$; 10,000 permutations).

Accessible Chromatin Regions Exhibit Signatures of Selection

An excess of substitutions on the Y chromosome could be caused by relaxed purifying selection or positive selection. To distinguish between these alternatives, we used short-read sequencing of 12 males from a freshwater population of stickleback fish (Shanfelter et al. 2019) to search for signatures of selection within ACRs. In the absence of selection, the ratio of polymorphism to sequence divergence should be equal between functional and nonfunctional sites. Deviation from this expectation would imply these regions are under selection. To explore this, we performed a modified McDonald–Kreitman (MK) test on noncoding sequences (Andolfatto 2005). We compared the number of X- and Y-specific substitutions to the number of polymorphisms segregating within a population for ACRs and nonfunctional control sequences. Except for stratum 2 on the X chromosome, sex chromosome ACRs from the liver and the testis had a greater proportion of sites fixed by positive selection (α), compared to autosomal ACRs ($P < 0.001$; chi-square test) (Table 1). We also identified the direction of selection acting on ACRs by calculating the odds-ratio of the MK test, known as the neutrality index (NI) (Rand and Kann 1996). Liver ACRs of the sex chromosomes had an $NI < 1$ ($P < 0.001$; chi-square test), consistent with positive selection (Fig. 3; Table 1). The only exception we found was for one of the younger strata on the X chromosome (stratum 2), which had an elevated NI consistent with purifying selection. Despite having a mean divergence lower than intergenic sequence, we found signatures of positive selection on the X chromosome for strata 1 and 3. In these cases, the departure from neutrality was driven by the low number of polymorphisms within ACRs compared to intergenic sequences, rather than high divergence within ACRs (Table 1). In these cases, the reduced polymorphism may be due to recent positive selection (Hughes et al. 2008; Stoletzki and Eyre-Walker 2011). Each stratum on the Y chromosome had a clear signature of positive selection ($NI < 1$; $P < 0.001$). Compared to the X chromosome and the autosome, the Y chromosome liver ACRs had stronger signatures of positive selection, indicative of a nonoverlapping bootstrap interval. For the testis ACRs, NI was indistinguishable between the X and Y chromosome, and both sex chromosomes had nonoverlapping bootstrap intervals with the representative autosome, indicating stronger positive selection acting on sex-linked ACRs. Similar to X-linked ACRs in liver, we found that testis ACRs in

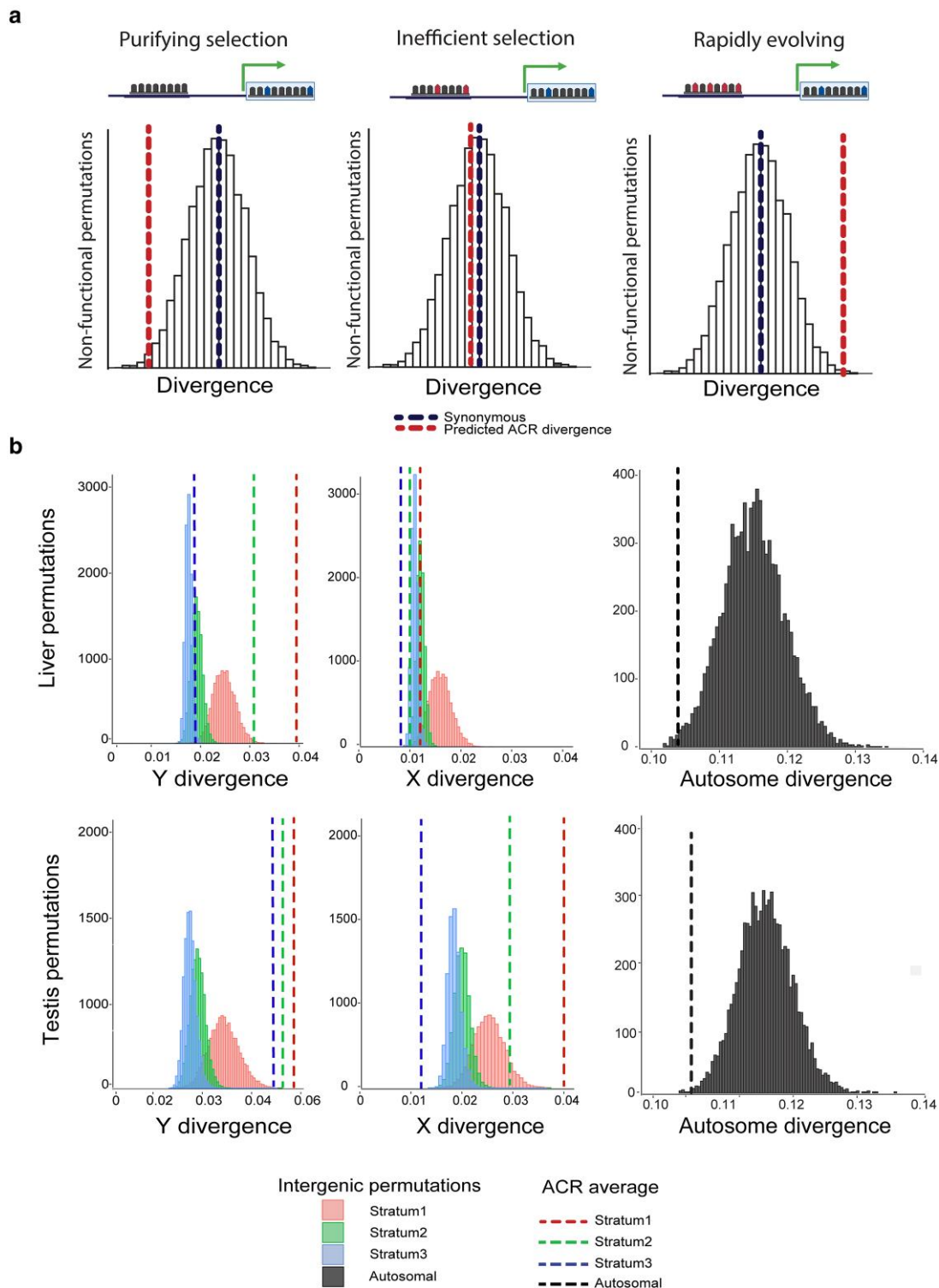


Fig. 2. Accessible chromatin divergence between threespine and ninespine stickleback. a) Predicted substitution rates within accessible regions. ACR substitution rates (red) lower than random nonfunctional control regions (white) and synonymous substitutions (blue) would indicate purifying selection (left). ACR substitution rates roughly equal to nonfunctional control regions would indicate mutations are accumulating due to inefficient selection from the loss of recombination (middle). ACR substitution rates higher than nonfunctional control regions would indicate positive selection. b) Y-linked ACR divergence is elevated over nonfunctional intergenic regions. X and Y ACR divergence was compared to randomly drawn nonfunctional regions matched for GC content (10,000 permutations). The average ACR divergence of each stratum is shown by the dotted line (stratum 1: pink, stratum 2: green, and stratum 3: blue). ACRs were identified on the X chromosome and mapped to the Y chromosome and homologous ninespine stickleback autosome 19 from liver (top) and testis (bottom). Autosome divergence is from chromosome 18. For the sex chromosomes, we identified variants that were unique to the X and Y, by aligning to the outgroup and ignoring fixed differences between species that likely occurred before the sex chromosomes evolved. X- and Y- specific variants within ACRs were summed and divided by the total length of the ACR to calculate divergence. Stratum 1: 442 liver ACRs and 30 testis ACRs; stratum 2: 228 liver ACRs and 37 testis ACRs; stratum 3: 129 liver ACRs and 46 testis ACRs; autosomal: 1056 liver ACRs and 305 testis ACRs. Graphic made with BioRender.com.

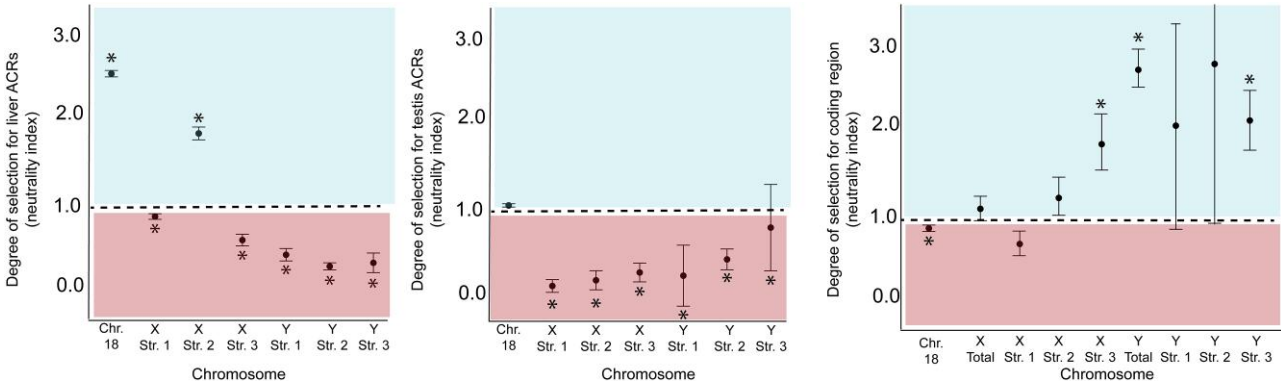


Fig. 3. Quantifying adaptive divergence and departure from neutrality among ACRs and coding regions. Sex-linked ACRs exhibited signatures of positive selection in both liver (left) and testis (middle). Sex-linked coding regions (right) are mostly under purifying selection or not significantly different than neutral expectations. The NI was calculated using the odds ratio of the McDonald Kreitman test. NI > 1 reflects purifying selection while NI < 1 is a signature of positive selection. Asterisks denote significant departures from neutrality determined by a chi-square test. Error bars represent 2 standard deviations calculated from 10,000 bootstrap replicates.

Table 1 MK test for noncoding ACRs

Chromosome	Stratum	Tissue	D(ACR)	D(INT)	P(ACR)	P(INT)	α	NI	Chi-square	P-value
Y	Stratum 1	Testis	739	555	18	46	0.706	0.294	20.77	$P < 0.001$
Y	Stratum 2	Testis	1292	912	2	10	0.859	0.141	8.65	$P = 0.003$
Y	Stratum 3	Testis	454	194	7	16	0.813	0.187	14.43	$P < 0.001$
Y	Stratum 1	Liver	89,313	52,716	243	923	0.845	0.155	870.81	$P < 0.001$
Y	Stratum 2	Liver	30,619	20,743	210	388	0.633	0.367	147.01	$P < 0.001$
Y	Stratum 3	Liver	16,889	16,875	321	412	0.222	0.778	11.13	$P < 0.001$
X	Stratum 1	Testis	878	760	452	18,109	0.978	0.022	6406.49	$P < 0.001$
X	Stratum 2	Testis	1225	1037	52	460	0.904	0.096	325.35	$P < 0.001$
X	Stratum 3	Testis	382	302	91	365	0.803	0.197	145.19	$P < 0.001$
X	Stratum 1	Liver	26,658	34,849	6009	10,003	0.215	0.785	176	$P < 0.001$
X	Stratum 2	Liver	7121	13,850	1265	1321	-0.863	1.863	224.76	$P < 0.001$
X	Stratum 3	Liver	9644	13,485	2618	7618	0.519	0.481	793.25	$P < 0.001$
Chr18	NA	Liver	23,380	25,607	9533	4193	-1.49	2.49	2028.33	$P < 0.001$
Chr18	NA	Testis	8715	9545	1509	1563	-0.057	1.057	2.05	$P = 0.153$

Table 2 MK test for coding regions

Chromosome	Subset of coding sites	DN	DS	PN	PS	NI	Chi-square	P-value
Y	Whole Chromosome	16,892	23,964	59	34	2.461	18.67	$P < 0.001$
X	Whole Chromosome	12,586	22,431	261	427	1.089	1.164	$P = 0.281$
Y	Stratum 1	3389	4143	12	7	2.090	2.526	$P = 0.112$
Y	Stratum 2	4240	7198	11	7	2.096	4.451	$P = 0.035$
Y	Stratum 3	8147	11,335	31	20	2.157	7.517	$P = 0.006$
X	Stratum 1	1901	3465	60	154	0.710	4.93	$P = 0.026$
X	Stratum 2	3549	6952	57	102	1.094	0.294	$P = 0.588$
X	Stratum 3	6245	10,776	135	140	1.664	17.88	$P < 0.001$
Chr18	Whole Chromosome	23,795	31,499	2581	3823	0.890	17.49	$P < 0.001$

stratum 3 on the X chromosome exhibited an NI < 1, due to a low number of polymorphisms in ACR rather than an excess in divergence.

We compared the patterns we observed across ACRs with coding regions. We found that coding regions across the entire Y chromosome had an excess of nonsynonymous polymorphisms (Fig. 3; NI > 1; $P < 0.001$; chi-square test), consistent with relaxed purifying selection on a degenerating Y chromosome. This pattern was largely driven by the youngest 2 strata of the Y chromosome (Table 2).

The oldest stratum (1) did not deviate from neutral expectations ($P = 0.112$). Coding regions across the X chromosome were not significantly different than neutral expectations (NI = 1; $P = 0.281$; chi-square test). When split by evolutionary strata only the youngest stratum (3) deviated from neutral expectations after correcting for multiple comparisons. In stratum 3, the coding regions exhibited purifying selection (NI > 1; $P_n < P_s$; $P < 0.001$; chi-square test). Unlike the Y chromosome, the entire X chromosome did not have an excess of nonsynonymous

polymorphism (Table 2). A gene-by-gene MK test revealed that 50% of X-linked gametologs had an NI > 1, compared to only 25% which had an excess of amino acid substitutions (supplementary material, Supplementary Material online). Importantly, sex-linked coding regions were not under widespread positive selection, like ACRs.

Reduced Polymorphism Throughout the Sex Chromosomes

Due to a lower effective population size, sex chromosomes are expected to have lower nucleotide diversity relative to the autosomes, given an equal sex ratio in the population. To test whether nucleotide diversity is reduced on the threespine stickleback sex chromosomes, we estimated average nucleotide diversity in sliding windows across the sex chromosomes and autosome 18. Across windows, we found Y chromosome nucleotide diversity was much lower than the expected 25% of nucleotide diversity compared to autosomes (Y chromosome π : 0.0001; autosome π : 0.0043; supplementary fig. S1, Supplementary Material online). This suggests that selection acting on linked deleterious or beneficial mutations has lowered nucleotide diversity beyond neutral expectations on the Y chromosome. The X chromosome is predicted to contain 75% of the nucleotide diversity compared to the autosomes. We found that the non-pseudoautosomal region (non-PAR) regions of the X chromosome had 45% of the amount of nucleotide diversity as autosomes, indicating the X chromosome may also be under selection, reducing nucleotide diversity beyond neutral expectations (non-PAR X chromosome π : 0.0019; autosome π : 0.0043; supplementary fig. S1, Supplementary Material online). We found that the Y chromosome had consistent signatures of low nucleotide diversity across all strata (supplementary fig. S1, Supplementary Material online). However, there were distinct differences in nucleotide diversity for each stratum and the PAR on the X chromosome (supplementary fig. S1, Supplementary Material online; $P < 0.001$ for all comparisons; Kruskal–Wallis and Dunn's test), suggesting each region of the X chromosome may be under different selection pressures depending on gene and regulatory content.

Y-Linked Substitutions are Correlated With Biased Expression From the X Chromosome

Selection should favor the loss of expression from the Y chromosome in order to silence coding regions that have accumulated deleterious mutations (Orr and Kim 1998; Bachtrog 2006; Lenormand et al. 2020). To test if Y-linked *cis*-regulatory divergence was associated with changes in expression, we compared ACR sequence divergence to changes in allelic expression on the X and Y from liver and testis RNA-seq transcriptomes (Peichel et al. 2020). To maximize the total number of genes for this analysis, we pooled genes across all 3 evolutionary strata. We found that sequence divergence among liver ACRs that were proximal to genes was a predictor of allele-specific expression. We observed higher expression of the X-linked

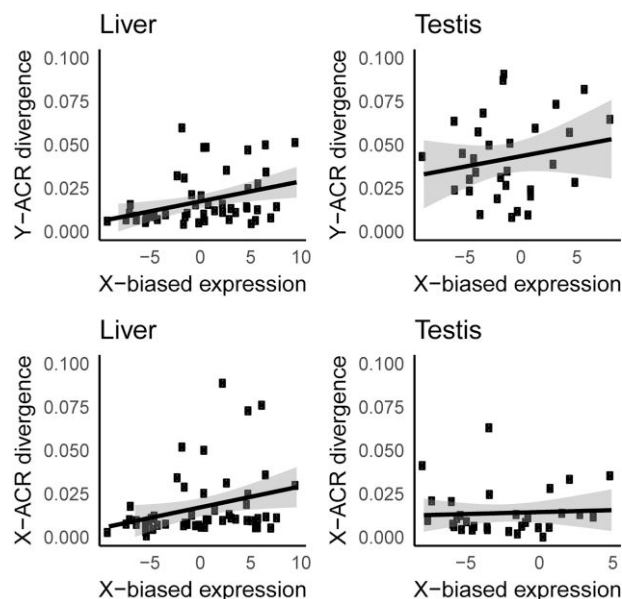


Fig. 4. Gametolog expression level compared to accessible chromatin region divergence. ACR divergence on the X and Y chromosomes is correlated with X-biased gene expression in both liver and testis tissues. RNA-Seq transcript counts that uniquely mapped to the X and Y chromosome were quantified to determine allele-specific expression for all gametologs. Gametolog expression was compared to the average divergence of Y-specific (a and b) and X-specific (c and d) mutations for ACRs within 50 kb of expressed genes. a) Y liver: $R = 0.40$, $P = 0.004$, $N = 51$; b) Y Testis: $R = 0.201$, $P = 0.26$, $N = 31$; c) X Liver: $R = 0.30$, $P = 0.033$, $N = 51$; d) X Testis: $R = 0.07$, $P = 0.94$, $N = 31$. Shaded regions represent 95% confidence intervals based on bootstrapping. The regressions were not significantly different between the X and Y chromosome for liver ($P = 0.590$, Fisher transformation) or testis ($P = 0.077$, Fisher transformation).

gametolog, relative to the Y-linked gametolog, when ACRs on the Y chromosome had more substitutions (Fig. 4A; $N = 51$; $P = 0.004$; $R = 0.4$; Spearman's rank correlation; 95% Confidence Interval via bootstrapping (0.1249, 0.5769)), indicating that ACR divergence explains around 16% of the X-biased expression observed in liver tissue. We also observed a correlation between expression and substitutions within ACRs on the X chromosome (Fig. 4C; $N = 51$; $P = 0.033$; $R = 0.30$; Spearman's rank correlation; 95% Confidence Interval from bootstrapping (0.0909, 0.4742)). The slopes of each linear regression were indistinguishable between the X and Y chromosome ($P = 0.590$; Fisher Z transformation). Although we do not have gene expression from the ninespine stickleback to determine which sex chromosome is deviating from the ancestral expression level, the Y-linked substitutions indicate the X-biased expression we observe is likely due to down-regulating the Y chromosome.

A similar trend was observed in testis tissue for Y-linked ACR divergence (Fig. 4), but the correlations were not significant (Fig. 4a; $N = 31$; $P = 0.26$; $R = 0.201$; Spearman's rank correlation; 95% Confidence Interval via bootstrapping (−0.0095, 0.5706)). There are fewer testis-expressed genes in this analysis. Therefore, the lack of significance could be due to sample size. To identify if sample size

Table 3 Gametologs with Y-biased expression in testes

Gene stable ID	Gene name	Gene description
ENSGACG00000003328	hipk3a	homeodomain interacting protein kinase 3a
ENSGACG00000004302	KCNA1	potassium voltage-gated channel, shaker-related subfamily, member 1a
ENSGACG00000006090	RFX7	regulatory factor X7b
ENSGACG00000006809	tdg.1	thymine DNA glycosylase, tandem duplicate 1
ENSGACG00000010054	hbp1	high mobility group-box transcription factor 1
ENSGACG00000010672	rora	retinoic acid receptor-related orphan receptor A, paralog a
ENSGACG00000010898	NA	NA
ENSGACG00000011130	mrpl23	mitochondrial ribosomal protein L23
ENSGACG00000011958	ergic2	si:ch211-225b10.3
ENSGACG00000012079	NA	lysine deficient protein kinase 1a
ENSGACG00000012796	NA	NA
ENSGACG00000012946	c2cd5	C2 calcium dependent domain containing 5
ENSGACG00000013060	IQSEC3	isoleucine and glutamine motif and Sec7 domain ArfGEF 3b
ENSGACG00000013333	SRGAP1	Rho GTPase activating protein 1a
ENSGACG00000013345	xylyt1	ribitol xylosyltransferase 1
ENSGACG00000013388	mtss1la	metastasis suppressor protein I-BAR domain containing 2a
ENSGACG00000013447	trabd	TraB domain containing
ENSGACG00000013796	CDH13	cadherin 13, H-cadherin (heart)
ENSGACG00000013862	nup93	nucleoporin 93
ENSGACG00000013972	cbfb	core-binding factor subunit beta
ENSGACG00000013994	hsf4	heat shock transcription factor 4
ENSGACG00000014090	NA	NA

had an effect, we performed a randomized down-sampling for the liver-expressed genes. Even with down-sampling, 68.0% of the permutations still produced significant correlations between ACR divergence and X-biased expression in the liver ($P < 0.05$; 10,000 permutations). This indicates the lack of correlation observed with the testis-expressed genes is likely not due to a smaller sample size. Despite having greater ACR divergence in testis compared to liver, 22 gametologs exhibited Y-biased expression in testis (Table 3). This suggests Y ACR evolution also contributes to up-regulation of male-specific genes.

Cis-regulatory Evolution is not Correlated With Deleterious Mutations in Coding Regions

The down-regulation of Y-linked gametologs could be adaptive if loss of expression follows the accumulation of deleterious coding substitutions (Orr and Kim 1998). We searched for evidence of adaptive silencing on the Y chromosome by comparing ACR divergence to the ratio of nonsynonymous to synonymous substitutions (d_N/d_S) within the coding region of each gene. An adaptive silencing model would be supported if Y-linked gametologs with an elevated d_N/d_S ratio also have elevated *cis*-regulatory divergence. Among genes shared between the X and Y chromosomes, we found no correlation between ACR nucleotide divergence and

d_N/d_S (supplementary fig. S2, Supplementary Material online; each stratum on X and Y: $P > 0.05$; Spearman's rank correlation), indicating coding regions with elevated ACR divergence are not more likely to have elevated substitutions in the amino acid sequence.

Elevation of d_N/d_S on the Y chromosome could indicate relaxed purifying selection or positive selection. Combining genes that are evolving under both could mask signatures of adaptive silencing. To account for this, we split gametologs on the Y chromosome into classes based on the signatures of selection from the MK test. We grouped gametologs into categories of purifying selection (NIHaldane > 0.25 ; $N = 262$), positive selection (NIHaldane < -0.25 ; $N = 41$), and genes evolving through relaxed purifying selection ($-0.25 > \text{NIHaldane} > 0.25$; $N = 219$). If selection was acting on ACRs to downregulate genes with deleterious mutations, we would expect to see a positive correlation between ACR divergence and d_N/d_S for genes with signatures of relaxed purifying selection. In contrast to this prediction, we found a weak negative correlation between ACR divergence and d_N/d_S for expressed genes under relaxed purifying selection (supplementary fig. S3, Supplementary Material online; $R = -0.293$; $P = 0.0155$; $N = 35$; Spearman's rank correlation). These genes also did not have more X-biased expression than other genes (supplementary fig. S4, Supplementary Material online; $P > 0.05$; Kruskal–Wallis test). Interestingly, we did find a positive correlation between ACR divergence and d_N/d_S for genes evolving under positive selection (supplementary fig. S4, Supplementary Material online; $R = 0.92$; $P = 0.002$; $N = 10$; Spearman's rank correlation). Overall, most genes on the Y seem to be evolving independently from neighboring regulatory regions, except for a small set of genes under positive selection.

We also examined whether ACR divergence was higher within genes that contained frameshift mutations, nonsense mutations, or in-frame deletions. These types of mutations are more likely to produce a nonfunctional or suboptimal peptide. We found no significant difference in the number of ACR substitutions from liver and testis tissue for coding regions that contained frameshift, nonsense mutations, or deletions, compared to coding regions that did not have these putatively deleterious mutations (functional) (supplementary fig. S5, Supplementary Material online) (all pairwise comparisons $P > 0.05$; Mann–Whitney U test). Overall, we found no evidence of *cis*-regulatory divergence associated with deleterious coding sequence divergence, suggesting the adaptive silencing model is unlikely the major driver of *cis*-regulatory evolution on the threespine stickleback sex chromosomes.

Discussion

The evolution of gene regulation on Y chromosomes has largely been studied in the context of RNA expression levels of gametologs (Muyle et al. 2012, 2018; White et al. 2015; Beaudry et al. 2017; Martin et al. 2019; Veltsos et al. 2019). Over time, gametolog expression is generally lost across most genes on the Y chromosome. With the

growing number of Y chromosome reference assemblies and sequencing methods to quickly profile functional *cis*-regulatory regions across the genome, it has become feasible to explore the molecular evolution of *cis*-regulatory regions that have led to altered expression patterns. Here, we observed elevated nucleotide substitutions and reduced polymorphisms within *cis*-regulatory elements across the threespine stickleback Y chromosome, consistent with positive selection driving this process. We found that increased rates of *cis*-regulatory divergence were associated with X-biased expression of gametologs, suggesting loss of expression from the Y chromosome. Without liver and testis gene expression from the nine-spine stickleback as an ancestral comparison, we were unable to confirm that X-biased gametolog expression is a result of down-regulation of the Y chromosome or up-regulation of the X chromosome. However, previous sequencing of a brain transcriptome showed expression patterns consistent with loss of Y expression rather than gain of X expression (White et al. 2015). It is therefore likely that the X-biased expression we observed in liver and testis tissue also occurs through the loss of Y chromosome expression, similar to what has been observed in other species with degenerating Y chromosomes (Muyle et al. 2012; Beaudry et al. 2017; Wei and Bachtrog 2019). Additional work will be necessary to identify the functional variants responsible for the X-biased expression pattern. We currently cannot determine whether the X-biased expression pattern is due to the accumulation of deleterious mutations from inefficient selection, adaptive substitutions that are accumulating in the ACRs, or a combination of the 2.

Several theoretical models of sex chromosome evolution have been developed that predict how gene expression evolves over time. Like coding regions, *cis*-regulatory regions will accumulate deleterious mutations because of reduced efficacy of purifying selection (Charlesworth and Charlesworth 2000; Bachtrog 2008). Without the removal of deleterious mutations, substitution rates within regulatory and coding regions approach the rates observed among nonfunctional intergenic sites. Theory predicts that substitution rates within regulatory regions can be elevated further through positive selection favoring beneficial mutations. The adaptive silencing model (Orr and Kim 1998) posits that Y gametologs that have accumulated maladaptive coding substitutions would be favored to be downregulated through *cis*-regulatory mutations. Only the functional X-linked gametolog would be expressed in males. An alternative model suggests that deleterious mutations accumulate in *cis*-regulatory elements throughout the nonrecombining region, initially downregulating all Y-linked gametologs (i.e. degeneration by regulatory evolution (Lenormand et al. 2020; Lenormand and Roze 2022)). As deleterious mutations simultaneously accumulate within coding regions, a positive-feedback loop ensues, leading to a higher substitution rate within *cis*-regulatory elements to selectively downregulate increasingly maladaptive coding regions. Our survey of ACR evolution in threespine stickleback fish allowed us to compare with each of the previously discussed theories.

Evidence for adaptive silencing (Orr and Kim 1998) would be present if alleles with elevated rates of coding divergence between the X and Y chromosomes were more likely silenced on the Y chromosome. Our results did not support this model. We found that the accumulation of X or Y ACR divergence had no association with relaxed purifying selection in nearby coding regions. Our results complement previous findings that have compared coding sequence evolution to gametolog expression. The level of expression from the Y chromosome is often not correlated with the overall number of deleterious mutations within coding regions (Bachtrog 2006; Bachtrog et al. 2008; Beaudry et al. 2017). While an elevation of d_N/d_S has been previously used as a proxy for the accumulation of deleterious variants on the Y, this association could also be confounded by positive selection. Here, we showed the lack of an association still held when only looking at genes evolving under relaxed selection. Additionally, we found no correlation between divergence within an ACR and the functional state of a coding region (i.e. containing frameshift or nonsense mutations). We found *cis*-regulatory evolution was correlated with expression differences, but that ACR divergence accumulated independent of coding divergence. A correlation between nucleotide substitutions in *cis*-regulatory regions and deleterious mutations in coding regions may be observed if ACRs and gene expression are profiled from additional tissues. In addition, we assigned ACRs to coding sequences based on overall proximity to a gene. Although this method of annotating ACRs is commonly used to assign *cis*-regulatory regions to genes (Connelly et al. 2014; Alexandre et al. 2018; Ricci et al. 2019), it is possible that some ACRs interact with other genes through long-range interactions based on chromatin configuration (Mifsud et al. 2015; Schoenfelder and Fraser 2019), reducing our ability to detect an association between deleterious mutations in coding regions and nucleotide substitutions in ACRs. Also, the number of substitutions within an ACR may not be a fully accurate predictor of the overall ability to silence a given gametolog on the Y chromosome. If only a small number of substitutions are needed to ablate transcription factor binding within an ACR, we would not expect to see a strong correlation between the number of deleterious mutations within coding sequences and the number of substitutions within ACRs. Additional functional work will be necessary to explore these alternatives.

Elevated substitution rates in *cis*-regulatory regions quickly following recombination suppression would be consistent with the model of degeneration by regulatory evolution (Lenormand et al. 2020; Lenormand and Roze 2022). Silencing of Y-linked alleles occurs first in this model, followed by the accumulation of deleterious mutations within coding regions. We found that putative *cis*-regulatory elements exhibited signatures of positive selection on the Y chromosome across all 3 evolutionary strata, including the 2 youngest strata that still contain most of the ancestral gene content (Peichel et al. 2020). The degeneration by regulatory evolution model predicts that the substitution rate within *cis*-regulatory elements

will accelerate once deleterious mutations begin to accumulate within coding regions. This will further suppress gametolog expression from the Y chromosome, masking maladaptive coding mutations. Consistent with this model, our results show *cis*-regulatory regions have stronger signatures of positive selection compared to autosomal ACRs, as well as both coding and noncoding regions on the sex chromosomes. This suggests there is selection to downregulate Y-linked alleles in liver. Alternatively, substitutions in ACRs could reflect adaptation to upregulate genes on the Y chromosome, to maintain expression, or to gain male-beneficial expression (Skaletsky et al. 2003; Martínez-Pacheco et al. 2020; Shaw and White 2022). Non-adaptive explanations should also be carefully considered as alternatives to adaptive models. For example, elevated substitution rates could be observed in Y-linked ACRs if they have higher mutation rates compared to synonymous and silent intergenic sites. Indeed, it has been suggested that male-biased mutation rates, due to increased cell division (Link et al. 2017), or specialized DNA repair pathways (Chang et al. 2022), could shape the evolution of the Y chromosome. These mutations could be fixed at a faster rate within populations and produce false adaptive signatures because of the smaller effective population size of the Y chromosome. However, the threespine stickleback fish Y chromosome does not show evidence of male-biased mutation rate at synonymous sites (White et al. 2015), and thus it remains unclear how these mutation biases would increase substitution rates uniquely at ACRs.

The degeneration by regulatory evolution model also indicates dosage compensation should evolve simultaneously for genes that are under strong stabilizing selection to maintain expression levels (Lenormand et al. 2020). Degeneration of *cis*-regulatory elements on the Y chromosome and loss of expression should select for up-regulation of the gametolog on the X chromosome to compensate for dosage loss. In this scenario, nucleotide substitutions should accumulate within ACRs on the X chromosome. We found that X-linked ACRs had a depletion of polymorphism, consistent with positive selection. Unlike Y-linked ACRs, the X-linked ACRs had low divergence over longer evolutionary timespans between ninespine and threespine stickleback fish. This suggests that selection has acted on some X-linked sites more recently within threespine stickleback populations to fix beneficial regulatory mutations. This distinction is consistent with Y-linked ACR divergence accumulating first, as predicted by the degeneration by regulatory evolution model (Lenormand et al. 2020). Selection on X-linked *cis*-regulatory elements within ACRs may still be ongoing. A similar finding was observed on the neo-sex chromosomes of *Drosophila miranda*, where the neo-X chromosome was enriched for signatures of selective sweeps, while the ancestral X chromosome region was not (Bachtrog et al. 2009).

We found the strongest signatures of positive selection on the X chromosome were from testis ACRs. Testis ACRs may have faster rates of sequence divergence on both the X and Y chromosome, from processes similar to

the faster X-effect (Charlesworth et al. 1987; Vicoso and Charlesworth 2009). There is widespread evidence for accelerated rates of coding evolution on X chromosomes compared to autosomes, especially in male-specific tissues (Baines et al. 2008; Meisel and Connallon 2013; Parsch and Ellegren 2013; Larson, et al. 2016; Kopania et al. 2022), as the X-linked mutations in hemizygous regions are always exposed to selection in males. Expression of X-linked gametologs may also evolve at faster rates (Khaitovich et al. 2005; Brawand et al. 2011; Meisel et al. 2012b; Coolon et al. 2015; Kopania et al. 2022), presumably due to mutations in *cis*-regulatory regions. The elevated substitution rate we observed on the X chromosome in testis ACRs may reflect male-beneficial selection to alter expression levels that are not necessarily connected to dosage compensation.

While we observed molecular signatures of positive selection within ACRs, we did not find sufficient evidence to suggest selection is associated with elevated X expression, as predicted by the degeneration by regulatory evolution model. While selection for dosage compensation could be ongoing, previous findings suggest that most genes on the threespine stickleback Y chromosome can be lost without consequence. Within the oldest stratum, over half of the Y-linked gametologs have been completely lost and males exhibit half expression, relative to females (White et al. 2015). Even across species with chromosome-wide dosage compensation mechanisms, many genes are not compensated completely (Cotton et al. 2013; Tukiainen et al. 2017), which suggests that loss of Y expression is not always deleterious. The dosage sensitivity of each sex-linked gene may be essential for predicting how rapidly *cis*-regulatory regions evolve on both sex chromosomes. In the degeneration by regulatory evolution model, if stabilizing selection is relaxed, chromosome-wide degeneration in *cis*-regulatory regions still occurs, the genes just do not evolve dosage compensation (Lenormand et al. 2020). Our results provide empirical support that *cis*-regulatory degeneration can occur without dosage compensation.

The expression balance of dosage-sensitive genes can also be maintained if the Y-linked gametolog does not degenerate. The coding sequence of haploinsufficient genes on Y chromosomes have been shown to be maintained through purifying selection. These genes presumably cannot be lost from the Y chromosome (Bellott et al. 2014, 2017; Peichel et al. 2020; Bellott and Page 2021). For these genes, if Y-linked *cis*-regulatory elements accumulated deleterious mutations, ancestral expression patterns could be maintained if compensatory transcription binding sites evolved. This would also result in signatures of increased substitution rates within ACRs on the Y chromosome, relative to intergenic regions. This type of compensatory evolution has been proposed for functionally critical genes in mammals (Chaix et al. 2008; Vermunt et al. 2016), and *Drosophila* (Landry et al. 2005; Arnold et al. 2014; Signor and Nuzhdin 2018). Functional analysis of what transcription binding sites are affected by Y-linked mutations could help identify which mutations are leading to

down-regulation of gametologs compared to those that maintain ancestral expression.

Recent models indicate inversions that suppress recombination between sex chromosomes can be selected to be retained within populations if sex-specific *trans*-acting regulators evolve to maintain optimal expression in both sexes (i.e. dosage compensation) (Lenormand and Roze 2022). This is based on the idea that if divergent X- and Y-linked *cis*-regulators were to recombine, it would lead to maladaptive expression levels in males or females. The evolutionary strata on the threespine stickleback Y chromosome evolved from at least 3 nested inversions that suppressed recombination (Peichel et al. 2020). Unlike this recent model, our results suggest these inversions were likely not under strong selection to be retained initially because of early evolution of dosage compensation. Chromosome-wide dosage compensation does not occur in the threespine stickleback (White et al. 2015) and we did not observe a corresponding evolution of X-linked up-regulation. Interestingly, the excess of X-linked mutations observed in testis ACRs may be reflective of a greater abundance of rapidly evolving sex-specific *trans*-regulators in the testes. However, additional work will be needed to identify the role that autosomal *trans*-regulators play in modifying gene expression of sex chromosomes in each sex, to test other aspects of the degeneration by regulatory evolution model.

Although we focused on the degeneration of *cis*-regulatory elements on the Y chromosome, it is important to note that Y chromosomes become masculinized over time and *cis*-regulatory elements may also be under positive selection for male-specific neo- or subfunctionalization (Soh et al. 2014; Mahajan and Bachtrog 2017; Bachtrog et al. 2019; Martínez-Pacheco et al. 2020; Peichel et al. 2020; Chang et al. 2022). Some of the ACRs we found with high divergence may actually reflect functional *cis*-regulatory regions associated with testis-specific genes. Interestingly, we found signatures of positive selection in ACRs near testis-expressed genes. Some of these genes were enriched for Y-biased expression. These findings provide evidence that ancestral gametolog expression levels can also change presumably through gain of new testis-related functions. Additional work will be necessary to identify the novel function these regulatory regions provide during spermatogenesis. An increase in testis expression was also found for ancestral gametologs in a comparative study across 17 species of mammals (Martínez-Pacheco et al. 2020). In these species, most Y-linked gametologs gained novel testis expression patterns, compared to their X-linked alleles. Rapid sex-linked *cis*-regulatory evolution may therefore be a universal phenomenon across species.

Y chromosomes often have much lower diversity overall, relative to neutral expectations (Lawson Handley et al. 2006; Wilson Sayres et al. 2014). This pattern has been attributed to purifying selection removing deleterious mutations and the linked neutral variation throughout the nonrecombining region (Wilson Sayres et al. 2014; Wilson Sayres 2018). Our findings revealed that nucleotide diversity on the threespine stickleback Y chromosome is also

much lower than neutral expectations. Our results suggest positive selection within ACRs could also be an important driver of reducing nucleotide diversity on the Y chromosome within populations. One important consideration is whether sex-biased demography may affect Y-linked diversity. If males are more variable in reproductive success than females, this could lower the expected effective population size of the Y chromosome, reducing nucleotide diversity. Little is known about the operating sex-ratios of stickleback fish. Some populations exhibit sex ratios that are female-biased (Rollins et al. 2017). This could contribute to the low nucleotide diversity we observed. However, simulations in human populations revealed that even drastic shifts in sex ratio could not entirely explain the low diversity observed on the Y chromosome (Wilson Sayres et al. 2014).

Conclusion

Together, our results provide evidence of positive selection driving accelerated rates of nucleotide substitution in *cis*-regulatory elements. Signatures of positive selection, even in the youngest 2 strata, indicate that *cis*-regulatory evolution can proceed rapidly following the suppression of recombination, leading to reduced gene expression from the Y chromosome. These results support some aspects of recent models of sex chromosome evolution, where *cis*-regulatory degeneration and silencing occur first on the Y chromosome. However, we found this can occur in the absence of dosage compensation, contrasting other assumptions of these models. Improvements in functional annotations of regulatory regions as well as an ever-growing collection of high-quality Y and W assemblies will allow continued empirical testing of new regulatory models of sex chromosome evolution.

Materials and Methods

Molecular Evolution of Accessible Chromatin Regions and Coding Regions

We used accessible chromatin regions (ACRs) from 2 different tissues. Liver ACRs were previously identified from 2 replicates (Naftaly et al. 2021) (NCBI BioProject PRJNA667175). We also collected testes from 2 juvenile males (~4.4 cm in standard length) of laboratory-reared threespine stickleback fish, originally isolated from Lake Washington (Seattle, Washington, USA) (NCBI BioProject PRJNA686097). The testis cells were immediately dissociated through homogenization in 1X phosphate-buffered saline (PBS) containing proteinase inhibitor cocktail (PIC, cOmplete tablets Roche) (PBS + PIC). The cells were fixed with 16% formaldehyde and washed twice with PBS + PIC, followed by lysis in 1 M Tris-HCl, pH = 8, 0.5 M EDTA, 10% NP-40, 50% glycerol/molecular grade H₂O, and 1X PIC. Nuclei were stained with 4',6-diamidino-2-phenylindole and counted with a hemocytometer. We diluted samples to 60,000 to 80,000 nuclei. ATAC-seq library preparation was conducted using previously established protocols (Lu et al. 2017; Naftaly et al. 2021). The libraries were

sequenced on an Illumina NextSeq (2 × 150 bp; Georgia Genomics and Bioinformatics Core). Reads were trimmed with Trimmomatic (v. 0.36) (Bolger et al. 2014), using a sliding window of 4 bases, trimming the remainder of the read when the average quality within a window dropped below 20. Residual adapter sequences were removed using Trimmomatic ILLUMINACLIP. Reads were filtered for a minimum length using MINLEN:30. We aligned the trimmed reads using Bowtie2 (v. 2.4.1) with default parameters (Langmead and Salzberg 2012). We filtered for alignments with a mapping quality greater than 20 using SAMtools (v1.14). We also removed polymerase chain reaction (PCR) duplicates by using the MarkDuplicates function in Picard (<https://github.com/broadinstitute/picard>).

We used a Tn5 control sample to normalize ATAC-seq reads to remove the effect of Tn5 bias. For the control, genomic DNA was extracted from a caudal fin clip of 1 male fish using a standard phenol-chloroform extraction. The ATAC-seq library preparation, whole-genome sequencing library preparation, and sequencing (Illumina HiSeq 2 × 150bp) of the control sample were completed by GENEWIZ (New Jersey, USA). Whole-genome sequencing was conducted in order to show the DNA sample was of sufficient quality to construct the Tn5 bias control. For the whole-genome sequencing, we recovered over 232 million reads with an average quality score of 38.59, indicating a high-quality sample. We trimmed residual adapters and low-quality sequences using Trimmomatic as previously described. Trimmed reads were aligned to the threespine stickleback genome using Bowtie2. The read coverage per base pair was calculated using BEDTools (v2.29, -d) (Quinlan and Hall 2010). The whole-genome sequencing sample had an average read depth of 90× where only ~10% of the genome was supported with <10 reads. Within these regions, only 6% had zero reads per base pair, indicating 94% of the genome could be queried for biased integration of Tn5. The Tn5 bias control produced over 177 million reads. The reads were trimmed from these reads using Trimmomatic with the same parameters. The trimmed reads were aligned to the threespine stickleback genome using Bowtie2 with default parameters. In order to use previous alignment coordinates between the X and Y chromosomes, we used the v.4 genome build of the X chromosome (Peichel et al. 2020). We used the v.5 genome build of chromosome 18 (Nath et al. 2021). Reads mapping to the mitochondria and unscaffolded regions were removed. PCR duplicates were removed using MarkDuplicates from Picard.

We evaluated concordance between replicates in several ways. We first searched for enrichment of ATAC-seq reads around the transcription start sites (TSSs) of expressed genes. Iso-Seq long-read sequencing was previously conducted in threespine stickleback fish to curate accurate transcription start site annotations across multiple tissues (Naftaly et al. 2021). We used deepTools (v3.5.1) computeMatrix reference-point (Ramírez et al. 2016) on the ATAC-seq alignments to assay read depth 3 kb around the complete set of annotated TSSs. We plotted enrichment of ATAC-seq reads around TSSs and compared with RNA-seq expression using

deepTools plotHeatmap with default settings, except that we used the -sortUsing=samples to sort the regions by expression. We found strong enrichment around TSSs, across both replicates for the same set of expressed genes (supplementary fig. S6, Supplementary Material online). ACRs were called for each replicate using MACS2 (v2.2.7.1) callpeak with the -keep-dup all parameter, and read depth was normalized with Tn5 control sequencing with the -c parameter (Zhang et al. 2008). The majority of ACRs were shared across replicates (liver: 67% and testis: 61%). We then used the *P*-value of each MACS2 peak to calculate the Irreproducibility Discovery Rate (IDR) (https://github.com/daniel-shaw1/Regulatory_divergence_paper/blob/main/Supplemental/IDR.r), (supplementary fig. S7, Supplementary Material online). The IDR determines the probability that a peak is reproducible between replicates, due to chance (Qunhua et al. 2011; Landt et al. 2012). Replicates that have IDR < 0.05 are generally considered to be highly reproducible. However, the IDR test is considered very stringent (Jalili et al. 2015), potentially removing many true regions of increased accessibility. Because both replicates exhibited high concordance, we pooled the aligned reads from both replicates and called peaks from the joint alignment using MACS2 as previously described, as pooling raw reads has been previously reported (Reske et al. 2020; Yan et al. 2020) as a strategy to find high-quality enriched regions. We found that most ACRs (liver: 79%, testis: 75%) from the pooled sample overlapped with low-scoring ACRs from the individual replicates (IDR < 0.05). This suggested a majority of ACRs in the pooled dataset were of high-quality in the individual replicates.

To define an ancestral set of ACRs, we mapped the nucleotide sequence of each noncoding ACR on the X chromosome to the homologous region on the Y chromosome using previously generated alignment coordinates (Peichel et al. 2020). A similar approach has been used previously to define the ancestral set of coding regions, aligning X-linked gene annotations to the Y chromosome (Peichel et al. 2020). We also attempted to identify Y-linked ACRs and calculate divergence in a similar method as we identified with X-linked ACRs. However, we found that MACS2 identified significantly fewer enriched regions of accessible chromatin on the Y chromosome compared to the X chromosome. The few ACRs that were identified also aligned to multiple regions on the X chromosome, suggesting these regions were repetitive. Using the X chromosome ACRs, we identified orthologous regions between the X chromosome, Y chromosome, and orthologous autosome (chromosome 19) from the ninespine stickleback fish (*Pungitius pungitius*; (Varadharajan et al. 2019) using BLAST+ (blastn v. 2.11.0) with default blastn parameters and -perc_identity set to 75. (Camacho et al. 2009). To ensure high-quality alignments, we filtered for uniquely mapping alignments that also had a bit-score >100. We used BLAST+ alignments to extract ACRs and the proximal 50 bp upstream and downstream, to create multiple sequence alignments for downstream analysis. We aligned the threespine stickleback X and Y ACR sequences to the

ninespine stickleback autosome sequence using MUSCLE (v 3.8.1551) (Edgar 2004) with default parameters. We used MUSCLE to create a multiple sequence alignment between the X sequence, Y sequence, and the sequence from the orthologous autosome in order to identify X- and Y- specific variants. We called single nucleotide variants using snp-sites (v2.5.1) (Page et al. 2016) with default parameters. We calculated the reported divergence rate by dividing the number of variants by the number of the aligned sites.

We compared ACR divergence on the sex chromosomes to ACR divergence on autosome 18, a chromosome similar in length to the X chromosome. We identified orthologous regions between the threespine stickleback autosome 18 and autosome 18 from the ninepine stickleback fish using BLAST+ (blastn v. 2.11.0) (Camacho et al. 2009) as previously described. We called variants between autosomal regions in a similar manner as the sex chromosomes, by creating a pairwise alignment using MUSCLE (v 3.8.1551) with default parameters and called single nucleotide variants using the same python script as above. We calculated a divergence rate by dividing the number of single nucleotide variants by the size of the aligned region.

Divergence within ACRs was compared with the divergence of coding regions of neighboring genes. We identified the closest gene to each ACR by running annotatepeaks.pl in homer (v 4.11) (Heinz et al. 2010) using default settings. Per default settings, ACRs were assigned to the closest gene within 50 kb of the TSS. We verified the homer peak annotations with a custom Python script that identifies ACRs near the TSS of genes (https://github.com/daniel-shaw1/Regulatory_divergence_paper/blob/main/Supplemental/Peaksneargenes.py). Both methods were 100% concordant with each other. We further verified a subset of the annotations (200 peaks total) by visualizing the ACRs in the threespine stickleback JBrowse genome browser (<https://stickleback.genetics.uga.edu/>). We found that 100% of the 200 X-linked ACRs were located proximal to the predicted annotated gene (supplementary fig. S8, Supplementary Material online). For coding divergence comparisons, we used previously reported estimates of synonymous (d_s) and nonsynonymous (d_N) divergence of gametologs between the X and Y chromosomes (Peichel et al. 2020). Coding regions with d_N/d_s ratios of 99 were omitted as these represent alignments with very little sequence divergence to estimate d_N/d_s accurately.

Estimating a Control Substitution Rate

To generate control regions, we randomly sampled non-functional intergenic regions throughout the X chromosome, Y chromosome, or chromosome 18 to estimate a neutral substitution rate. Regulatory regions tend to be GC-rich, which can be prone to higher mutation rates through mechanisms like GC-biased gene conversion or spontaneous deamination of methylated cytosines (Nesta et al. 2021). We therefore GC-matched the randomly drawn intergenic regions with the GC content of ACRs. We calculated the GC percentage of ACRs using a modified Perl script (countbp.pl, Nicholas Navin, <http://www.navinlab.com/>

bioperl/bioperl/gc_content.html). For each ACR, a random intergenic region was drawn from the X chromosome equal in length and with a GC content within 2.0%. Intergenic regions were defined as any region that fell outside of annotated functional regions from a combination of Ensembl annotations (Cunningham et al. 2022), Isoseq transcripts (Naftaly et al. 2021), Y chromosome annotations (Peichel et al. 2020), repetitive elements (Peichel et al. 2020; Nath et al. 2021) and ACRs from this study. This was performed for each stratum individually using bedtools shuffle -l ACRs.bed -incl XYalignment.bed -excl annotations.bed. We extracted 10,000 sets of GC-matched intergenic regions from each of the 3 evolutionary strata on the threespine stickleback sex chromosomes (set size same as number of Liver ACRs: stratum 1: 442, stratum 2: 228, stratum 3: 199; Testis ACRs: stratum 1: 30, stratum 2: 37, stratum 3: 51). Due to a limited number of intergenic sites shared between the sex chromosomes, we sampled with replacement to align a sufficient number of intergenic regions. We identified the orthologous regions to the ninepine genome assembly using BLAST+ (blastn v2.11.0) (Camacho et al. 2009). We filtered for alignments that mapped uniquely to chromosome 19 in the ninepine stickleback fish, had an alignment length equal to 75% of the threespine query sequence, and had a bit-score > 100. We generated multiple sequence alignments using MUSCLE (v 3.8.1551), for each intergenic permutation and calculated X- and Y-specific substitution rate as previously described.

Estimating Within Population Nucleotide Diversity

We used whole-genome short-read sequencing of 12 males from the Lake Washington population (Washington, USA; NCBI SRA SRP137809) to estimate nucleotide diversity across the sex chromosomes. The raw reads were trimmed with Trimmomatic (v. 0.39) (Bolger et al. 2014), using a sliding window of 4 bases, trimming the remainder of the read when the average quality within a window dropped < 15. The leading and trailing base pairs below quality 3 of every read were removed along with any residual adapter sequence. After trimming, any reads below a minimum length of 36 were discarded. We aligned the trimmed reads using Bowtie2 (v. 2.4.5) with default parameters (Langmead and Salzberg 2012). We also marked PCR duplicates by using the MarkDuplicates function in Picard (v. 2.26.10) (<https://github.com/broadinstitute/picard>). Single nucleotide polymorphism (SNP) genotyping was conducted using the GATK software package (v. 4.2.5.0). We called variants using HaplotypeCaller in genomic variant call format (GVCF) mode. We then joint-called variants using GenotypeGVCFs using a genomics database created by GenomicsDBImport.

We only considered biallelic SNPs in our estimates of nucleotide diversity and we filtered for high-quality genotypes using several different methods. On the sex chromosomes, sites should only have hemizygous genotypes in males in the noncrossover region outside of the pseudoautosomal region. Any sites that are heterozygous would be caused by errors in read alignment. We therefore did not consider

sites that were heterozygous in the noncrossover region on the X or Y chromosomes. We also filtered out sites that exhibited too low or high of read depth, which would be indicative of alignment errors. To do this, we did not consider sites that were less than one-half or greater than double the median read depth of each chromosome (X and Y chromosomes: positions were retained if the read depth was between 3.5 and 14; chromosome 18: positions were retained if the read depth was between 6.5 and 26). Read filtering was conducted using custom Perl scripts.

Nucleotide diversity (π) was estimated on a per-site basis ($-\text{site-}\pi$) across each chromosome using *vcftools* (v. 0.1.16) (Danecek et al. 2011). Nucleotide diversity was averaged within 10,000 bp nonoverlapping windows across chromosomes X, Y, and 18.

Testing for Positive Selection Using Population Polymorphism

We performed a modified version of the MK test (Andolfatto 2005) by treating divergence (D) and diversity (P) within ACRs as the nonsynonymous factor, and non-functional intergenic sequence previously generated was used as synonymous factor. As intergenic sequence is independent of each ACR, we pooled sites from each factor as previously described (Andolfatto 2005). We counted the total number of polymorphic sites or fixed substitutions (divergence) for each factor to calculate α , the fraction of substitutions that are likely fixed due to positive selection. Divergence was estimated either between the threespine stickleback X and Y chromosomes or between the threespine and ninespine stickleback autosome 18. Negative alpha values were interpreted as regions that have rates of adaptive evolution near zero (Good et al. 2013). We also used these counts to estimate the NI (Rand and Kann 1996). An NI >1 indicates an excess of ACR divergence within species, while an NI <1 indicates an excess of ACR divergence between species (or between the X and Y). An index <1 would be consistent with positive selection. To test for significant departures from neutrality within each stratum or chromosome we performed a chi-square test to compare the rates of d_N/d_S to p_N/p_S . Standard deviation was determined through non-parametric bootstrapping. We bootstrapped 10,000 replicates for each factor within the MK or NI test and computed both metrics for each replicate.

We also performed an MK test on coding regions. We aligned the coding sequence of all threespine stickleback X chromosome Ensembl predicted genes (build 95) to the Y chromosome and the ninespine stickleback outgroup (autosome 19; v. 7) (Varadharajan et al. 2019) using Exonerate (v. 2.4.0) (Slater and Birney 2005) with parameters: $-\text{model coding2genome} -M 2000 -\text{bestn} 1 -\text{annotation}$. The Ensembl coding regions from the threespine stickleback autosome 18 were also aligned to the ninespine stickleback autosome 18. The highest-scoring alignment from Exonerate was used for the MK test. Divergence and polymorphism were tallied at each position of the alignment. If a position was variable, we determined if this was a synonymous or nonsynonymous mutation in the threespine stickleback by

comparing it to the ninespine stickleback translated codon. Sites were omitted if they contained any N characters or contained any missing genotype data among the 12 males within the variant call format (VCF) file. Codons were omitted if they contained any gaps from the alignment. We only considered biallelic sites when counting polymorphisms and fixed substitutions within the Lake Washington population. We used custom Perl scripts to count the total number of divergent and polymorphic sites. To compare to ACRs, we calculated α and NI by pooling sites across all coding regions. For gene-by-gene MK tests, we log-transformed NI to include genes with a count of zero for any particular category of divergence or polymorphism and refer to the value as NI_{Haldane} (Haldane 1956; Stoletzki and Eyre-Walker 2011).

Comparison With Allele-specific Expression Patterns

We measured allele-specific expression of transcripts on the X and Y chromosome using 3 replicate liver samples (NCBI BioProject PRJNA591630) and 3 replicate testis samples (NCBI BioProject PRJNA591630). We trimmed the RNA-seq reads with Trimmomatic (v. 0.36) (Bolger et al. 2014) using a sliding window of 4 bases, trimming the remainder of the read when the average quality within a window dropped below 20 (SLIDINGWINDOW:4:20). Residual sequencing adapters were also removed using ILLUMINACLIP. We aligned RNA-seq reads using Tophat2 (v2.2.1) with default parameters (Kim et al. 2013). We filtered for alignments with a map quality greater than 25 using SAMtools (v1.14) (Li et al. 2009) to identify reads that map uniquely on the X- and Y-alleles. Similar map quality filters have been used to distinguish short reads between sex chromosomes in *Drosophila* and stickleback fish (Ellison and Bachtrog 2019; Peichel et al. 2020) and between sexes with heteromorphic sex chromosomes in guppies (Kirkpatrick et al. 2022). Read counts were obtained using htseq-count (v 0.9.1) (Anders et al. 2015). Default parameters were used with the addition of $-\text{stranded} = \text{no}$ and $-\text{nonunique} = \text{all}$ to maximize our ability to count reads on the Y chromosome. We used a custom gene transfer format (GTF) file for htseq-count that included previously determined start and end sites of all Ensembl predicted transcripts that are shared between the X and Y chromosomes (Peichel et al. 2020). Reads were counted on the X and Y across the entire transcript as a feature. Transcripts were removed from the analysis if they had an RNA expression count of 0 in all samples. For each tissue, we calculated the average read count for each gametolog across the 3 tissue replicates. We calculated the allele-specific expression as $\log_2$ (X transcript average/Y transcript average).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the National Science Foundation MCB 1943283 (M.A.W.), the National

Institutes of Health R01GM147312 (M.A.W.) and T32GM007103 (A.S.N.), and by the University of Georgia Research Foundation (D.E.S.). We would like to thank Bob Schmitz and Casey Bergman for helpful discussions related to the project.

Conflict of Interest

The authors declare no conflicts of interests.

Data Availability

The data for this manuscript is available within the short reads archive under accession numbers: PRJNA667175, PRJNA591630, PRJNA686097, SRP137809.

Scripts and bioinformatic pipelines are available at: https://github.com/daniel-shaw1/Regulatory_divergence_paper

References

- Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, Sullivan AM, Bemm F, Jolic D, Arsovski AA, *et al.* Complex relationships between chromatin accessibility, sequence divergence, and gene expression in *Arabidopsis thaliana*. *Mol Biol Evol.* 2018;**35**(4):837–854. <https://doi.org/10.1093/molbev/msx326>.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;**31**(2):166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
- Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature.* 2005;**437**(7062):1149–1152. <https://doi.org/10.1038/nature04107>.
- Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during *cis*-regulatory evolution. *Nat Genet.* 2014;**46**(7):685–692. <https://doi.org/10.1038/ng.3009>.
- Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* 2013;**14**(3):R26. <https://doi.org/10.1186/gb-2013-14-3-r26>.
- Bachtrog D. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. *Genome Res.* 2005;**15**(10):1393–1401. <https://doi.org/10.1101/gr.3543605>.
- Bachtrog D. Expression profile of a degenerating neo-Y chromosome in *Drosophila*. *Curr Biol.* 2006;**16**(17):1694–1699. <https://doi.org/10.1016/j.cub.2006.07.053>.
- Bachtrog D. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics.* 2008;**179**(3):1513–1525. <https://doi.org/10.1534/genetics.107.084012>.
- Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet.* 2013;**14**(2):113–124. <https://doi.org/10.1038/nrg3366>.
- Bachtrog D, Hom E, Wong KM, Maside X, De Jong P. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol.* 2008;**9**(2):R30. <https://doi.org/10.1186/gb-2008-9-2-r30>.
- Bachtrog D, Jensen JD, Zhang Z. Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* 2009;**7**(4):e1000082. <https://doi.org/10.1371/journal.pbio.1000082>.
- Bachtrog D, Mahajan S, Bracewell R. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nat Ecol Evol.* 2019;**3**(11):1587–1597. <https://doi.org/10.1038/s41559-019-1009-9>.
- Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, *et al.* Sex determination: why so many ways of doing it? *PLoS Biol.* 2014;**12**(7):e1001899. <https://doi.org/10.1371/journal.pbio.1001899>.
- Baines JF, Sawyer SA, Hartl DL, Parsch J. Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. *Mol Biol Evol.* 2008;**25**(8):1639–1650. <https://doi.org/10.1093/molbev/msn111>.
- Beaudry FEG, Barrett SCH, Wright SI. Genomic loss and silencing on the Y chromosomes of *Rumex*. *Genome Biol Evol.* 2017;**9**(12):3345–3355. <https://doi.org/10.1093/gbe/evx254>.
- Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghul S, Graves T, Rock S, *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature.* 2014;**508**(7497):494–499. <https://doi.org/10.1038/nature13206>.
- Bellott DW, Page DC. Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Res.* 2021;**31**(2):198–210. <https://doi.org/10.1101/gr.268516.120>.
- Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, Galkina S, Pyntikova T, Koutseva N, Graves T, *et al.* Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet.* 2017;**49**(3):387–394. <https://doi.org/10.1038/ng.3778>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 2014;**30**(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M. The evolution of gene expression levels in mammalian organs. *Nature.* 2011;**478**(7369):343–348. <https://doi.org/10.1038/nature10532>.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;**10**(12):1213–1218. <https://doi.org/10.1038/nmeth.2688>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;**10**(1):421. <https://doi.org/10.1186/1471-2105-10-421>.
- Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA. Evolution of primate gene expression: drift and corrective sweeps? *Genetics.* 2008;**180**(3):1379–1389. <https://doi.org/10.1534/genetics.108.089623>.
- Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuent AM. Unique structure and positive selection promote the rapid divergence of *Drosophila* Y chromosomes. *eLife.* 2022;**11**:e75795. <https://doi.org/10.7554/eLife.75795>.
- Charlesworth B. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci USA.* 1978;**75**(11):5618–5622. <https://doi.org/10.1073/pnas.75.11.5618>.
- Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 2000;**355**(1403):1563–1572. <https://doi.org/10.1098/rstb.2000.0717>.
- Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat.* 1987;**130**(1):113–146. <https://doi.org/10.1086/284701>.
- Connelly CF, Wakefield J, Akey JM. Evolution and genetic architecture of chromatin accessibility and function in yeast. *PLoS Genet.* 2014;**10**(7):e1004427. <https://doi.org/10.1371/journal.pgen.1004427>.
- Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the *Drosophila* X chromosome. *Mol Biol Evol.* 2015;**32**(10):2605–2615. <https://doi.org/10.1093/molbev/msv135>.
- Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 2013;**14**(11):R122. <https://doi.org/10.1186/gb-2013-14-11-r122>.
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean Irina M, Austine-Orimoloye O, Azov Andrey G, Barnes I, Bennett

- R, et al. Ensembl 2022. *Nucleic Acids Res.* 2022;**50**(D1):D988–D995. <https://doi.org/10.1093/nar/gkab1049>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;**27**(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;**32**(5):1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Ellison C, Bachtrög D. Recurrent gene co-amplification on Drosophila X and Y chromosomes. *PLoS Genet.* 2019;**15**(7):e1008251. <https://doi.org/10.1371/journal.pgen.1008251>.
- Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R, Fischer A, et al. Comparative population genomics of the ejaculate in humans and the Great Apes. *Mol Biol Evol.* 2013;**30**(4):964–976. <https://doi.org/10.1093/molbev/mst005>.
- Haldane J. The estimation and significance of the logarithm of a ratio of frequencies. *Ann Hum Genet.* 1956;**20**(4):309–311. <https://doi.org/10.1111/j.1469-1809.1955.tb01285.x>.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;**38**(4):576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- Horvath R, Josephs EB, Pesquet E, Stinchcombe JR, Wright SJ, Scofield D, Slotte T. Selection on accessible chromatin regions in *Capsella grandiflora*. *Mol Biol Evol.* 2021;**38**(12):5563–5575. <https://doi.org/10.1093/molbev/msab270>.
- Hughes AL, Friedman R, Rivailler P, French JO. Synonymous and non-synonymous polymorphisms versus divergences in bacterial genomes. *Mol Biol Evol.* 2008;**25**(10):2199–2209. <https://doi.org/10.1093/molbev/msn166>.
- Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki C, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature.* 2012;**483**(7387):82–86. <https://doi.org/10.1038/nature10843>.
- Hughes JF, Skaletsky H, Pyntikova T, Graves TA, Van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C, et al. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature.* 2010;**463**(7280):536–539. <https://doi.org/10.1038/nature08700>.
- Hughes JF, Skaletsky H, Pyntikova T, Koutseva N, Raudsepp T, Brown LG, Bellott DW, Cho T-J, Dugan-Rocha S, Khan Z, et al. Sequence analysis in *Bos taurus* reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Res.* 2020;**30**(12):1716–1726. <https://doi.org/10.1101/gr.269902.120>.
- Jalili V, Matteucci M, Masseroli M, Morelli MJ. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics.* 2015;**31**(17):2761–2769. <https://doi.org/10.1093/bioinformatics/btv293>.
- Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science.* 2005;**309**(5742):1850–1854. <https://doi.org/10.1126/science.1108296>.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;**14**(4):R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
- Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Scharf M. Evolution of the canonical sex chromosomes of the guppy and its relatives. *G3 (Bethesda).* 2022;**12**(2):jkab435. <https://doi.org/10.1093/g3journal/jkab435>.
- Kopania EEK, Larson EL, Callahan C, Keeble S, Good JM. Molecular evolution across mouse spermatogenesis. *Mol Biol Evol.* 2022;**39**(2):msac023. <https://doi.org/10.1093/molbev/msac023>.
- Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics.* 2005;**171**(4):1813–1822. <https://doi.org/10.1534/genetics.105.047449>.
- Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;**22**(9):1813–1831. <https://doi.org/10.1101/gr.136184.111>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;**9**(4):357–359. <https://doi.org/10.1038/nmeth.1923>.
- Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BA, Smith AD, Dean MD, Good JM. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics.* 2016;**203**(4):1841–1857. <https://doi.org/10.1534/genetics.116.186825>.
- Lawrie DS, Messer PW, Hershberg R, Petrov DA. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* 2013;**9**(5):e1003527. <https://doi.org/10.1371/journal.pgen.1003527>.
- Lawson Handley LJ, Berset-Brändli L, Perrin N. Disentangling reasons for low Y chromosome variation in the greater white-toothed shrew (*Crocidura russula*). *Genetics.* 2006;**173**(2):935–942. <https://doi.org/10.1534/genetics.105.050203>.
- Lenormand T, Fyon F, Sun E, Roze D. Sex chromosome degeneration by regulatory evolution. *Curr Biol.* 2020;**30**(15):3001–3006.e5. <https://doi.org/10.1016/j.cub.2020.05.052>.
- Lenormand T, Roze D. Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science.* 2022;**375**(6581):663–666. <https://doi.org/10.1126/science.abj1813>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;**25**(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Link V, Aguilar-Gómez D, Ramírez-Suástegui C, Hurst LD, Cortez D. Male mutation bias is the main force shaping chromosomal substitution rates in monotreme mammals. *Genome Biol Evol.* 2017;**9**(9):2198–2210. <https://doi.org/10.1093/gbe/evx155>.
- Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* 2017;**45**(6):e41–e41. <https://doi.org/10.1093/nar/gkw1179>.
- Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants.* 2019;**5**(12):1250–1259. <https://doi.org/10.1038/s41477-019-0548-z>.
- Mahajan S, Bachtrög D. Convergent evolution of Y chromosome gene content in flies. *Nat Commun.* 2017;**8**(1):785. <https://doi.org/10.1038/s41467-017-00653-x>.
- Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrög D. De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol.* 2018;**16**(7):e2006348. <https://doi.org/10.1371/journal.pbio.2006348>.
- Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GAB, Touzet P. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. *Genome Biol Evol.* 2019;**11**(2):350–361. <https://doi.org/10.1093/gbe/evz001>.
- Martínez-Pacheco M, Tenorio M, Almonte L, Fajardo V, Godínez A, Fernández D, Cornejo-Páramo P, Díaz-Barba K, Halbert J, Liechti A, et al. Expression evolution of ancestral XY gametologs across all Major groups of placental mammals. *Genome Biol Evol.* 2020;**12**(11):2015–2028. <https://doi.org/10.1093/gbe/evaa173>.
- Meisel RP, Connallon T. The faster-X effect: integrating theory and data. *Trends Genet.* 2013;**29**(9):537–544. <https://doi.org/10.1016/j.tig.2013.05.009>.
- Meisel RP, Malone JH, Clark AG. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res.* 2012a;**22**(7):1255–1265. <https://doi.org/10.1101/gr.132100.111>.

- Meisel RP, Malone JH, Clark AG. Faster-X evolution of gene expression in *Drosophila*. *PLoS Genet.* 2012b;**8**(10):e1003013. <https://doi.org/10.1371/journal.pgen.1003013>.
- Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015;**47**(6):598–606. <https://doi.org/10.1038/ng.3286>.
- Muller HJ. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics.* 1918;**3**(5):422–499. <https://doi.org/10.1093/genetics/3.5.422>.
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 2012;**10**(4):e1001308. <https://doi.org/10.1371/journal.pbio.1001308>.
- Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Hobza R, Picard F, Widmer A, *et al.* Genomic imprinting mediates dosage compensation in a young plant XY system. *Nat Plants.* 2018;**4**(9):677–680. <https://doi.org/10.1038/s41477-018-0221-y>.
- Naftaly AS, Pau S, White MA. Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Res.* 2021;**31**(8):1486–1497. <https://doi.org/10.1101/gr.274282.120>.
- Nath S, Shaw DE, White MA. Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 (Bethesda).* 2021;**11**(2):jkab007. <https://doi.org/10.1093/g3journal/jkab007>.
- Nesta AV, Tafur D, Beck CR. Hotspots of human mutation. *Trends Genet.* 2021;**37**(8):717–729. <https://doi.org/10.1016/j.tig.2020.10.003>.
- Orr HA, Kim Y. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics.* 1998;**150**(4):1693–1698. <https://doi.org/10.1093/genetics/150.4.1693>.
- Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics.* 2016;**2**(4):e000056. <https://doi.org/10.1099/mgen.0.000056>.
- Parsch J, Ellegren H. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 2013;**14**(2):83–87. <https://doi.org/10.1038/nrg3376>.
- Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM, Kingsley DM, *et al.* Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* 2020;**21**(1):177. <https://doi.org/10.1186/s13059-020-02097-x>.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;**26**(6):841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Qunhua L, James BB, Haiyan H, Peter JB. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat.* 2011;**5**:1752–1779. <https://doi.org/10.1214/11-AOAS466>.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;**44**(W1):W160–W165. <https://doi.org/10.1093/nar/gkw257>.
- Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol.* 1996;**13**(6):735–748. <https://doi.org/10.1093/oxfordjournals.molbev.a025634>.
- Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin.* 2020;**13**(1):22. <https://doi.org/10.1186/s13072-020-00342-y>.
- Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, *et al.* Widespread long-range *cis*-regulatory elements in the maize genome. *Nat Plants.* 2019;**5**(12):1237–1249. <https://doi.org/10.1038/s41477-019-0547-0>.
- Rodríguez Lorenzo JL, Hobza R, Vyskot B. DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *BMC Genomics.* 2018;**19**(1):540. <https://doi.org/10.1186/s12864-018-4936-y>.
- Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome—patterns and consequences. *Mol Ecol.* 2013;**22**(11):3014–3027. <https://doi.org/10.1111/mec.12322>.
- Rollins JL, Chiang P, Waite JN, Von Hippel FA, Bell MA. Jacks and Jills: alternative life-history phenotypes and skewed sex ratio in anadromous threespine stickleback (*Gasterosteus aculeatus*). *Evol Ecol Res.* 2017;**18**:363–382.
- Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. Heterogeneous histories of recombination suppression on stickleback sex chromosomes. *Mol Biol Evol.* 2021;**38**(10):4403–4418. <https://doi.org/10.1093/molbev/msab179>.
- Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet.* 2019;**20**(8):437–455. <https://doi.org/10.1038/s41576-019-0128-0>.
- Shanfelter AF, Archambeault SL, White MA. Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biol Evol.* 2019;**11**(6):1552–1572. <https://doi.org/10.1093/gbe/evz090>.
- Shaw DE, White MA. The evolution of gene regulation on sex chromosomes. *Trends Genet.* 2022;**38**(8):844–855. <https://doi.org/10.1016/j.tig.2022.04.004>.
- Signor SA, Nuzhdin SV. The evolution of gene expression in *cis* and *trans*. *Trends Genet.* 2018;**34**(7):532–544. <https://doi.org/10.1016/j.tig.2018.03.007>.
- Singh ND, Koerich LB, Carvalho AB, Clark AG. Positive and purifying selection on the *Drosophila* Y chromosome. *Mol Biol Evol.* 2014;**31**(10):2612–2623. <https://doi.org/10.1093/molbev/msu203>.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature.* 2003;**423**(6942):825–837. <https://doi.org/10.1038/nature01722>.
- Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005;**6**(1):31. <https://doi.org/10.1186/1471-2105-6-31>.
- Soh YQS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL, *et al.* Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell.* 2014;**159**(4):800–813. <https://doi.org/10.1016/j.cell.2014.09.052>.
- Stoletzki N, Eyre-Walker A. Estimation of the neutrality index. *Mol Biol Evol.* 2011;**28**(1):63–70. <https://doi.org/10.1093/molbev/msq249>.
- Tomaszkiewicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D, O'Brien PCM, Chikhi R, Ryder OA, *et al.* A time- and cost-effective strategy to sequence mammalian Y chromosomes: an application to the de novo assembly of gorilla Y. *Genome Res.* 2016;**26**(4):530–540. <https://doi.org/10.1101/gr.199448.115>.
- Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, *et al.* Landscape of X chromosome inactivation across human tissues. *Nature.* 2017;**550**(7675):244–248. <https://doi.org/10.1038/nature24265>.
- Varadarajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS, Merilä J. A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol Evol.* 2019;**11**(11):3291–3308. <https://doi.org/10.1093/gbe/evz240>.
- Veltsos P, Ridout KE, Toups MA, González-Martínez SC, Muyle A, Emery O, Rastas P, Hudzieczek V, Hobza R, Vyskot B, *et al.* Early sex-chromosome evolution in the diploid dioecious plant *Mercurialis annua*. *Genetics.* 2019;**212**(3):815–835. <https://doi.org/10.1534/genetics.119.302045>.
- Vermunt MW, Tan SC, Castelijn B, Geeven G, Reinink P, De Bruijn E, Kondova I, Persengiev S, Bontrop R, Cuppen E, *et al.* Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat Neurosci.* 2016;**19**(3):494–503. <https://doi.org/10.1038/nn.4229>.

- Vicoso B, Charlesworth B. Effective population size and the faster-X effect: an extended model. *Evolution*. 2009;**63**(9):2413–2426. <https://doi.org/10.1111/j.1558-5646.2009.00719.x>.
- Wei KHC, Bachtrog D. Ancestral male recombination in *Drosophila albomicans* produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. *PLoS Genet*. 2019;**15**(11):e1008502. <https://doi.org/10.1371/journal.pgen.1008502>.
- White MA, Kitano J, Peichel CL. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol Biol Evol*. 2015;**32**(8):1981–1995. <https://doi.org/10.1093/molbev/msv078>.
- Wilson Sayres MA. Genetic diversity on the sex chromosomes. *Genome Biol Evol*. 2018;**10**(4):1064–1078. <https://doi.org/10.1093/gbe/evy039>.
- Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced diversity on human Y chromosomes. *PLoS Genet*. 2014;**10**(1):e1004064. <https://doi.org/10.1371/journal.pgen.1004064>.
- Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol*. 2020;**21**(1):22. <https://doi.org/10.1186/s13059-020-1929-3>.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008;**9**(9):R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.