**Title: Positive selection drives *cis*-regulatory evolution across the threespine stickleback Y chromosome**

**Daniel E. Shaw, Alice Shanfelter Naftaly, and Michael A. White***

**Department of Genetics, University of Georgia, Athens, GA, 30602**

**Corresponding Author:**

**Michael A. White**

**Department of Genetics, University of Georgia**

**120 Green St,**

**Athens, GA 30602**

**whitem@uga.edu**

## 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 downregulation 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 three 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 non-functional 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.

## 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; White, et al. 2015; Bellott, et al. 2017; 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; Ayers, et al. 2013; Singh, et al. 2014; White, et al. 2015; Beaudry, et al. 2017; Muyle, et al. 2018; 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 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 non-coding 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; Bellott, et al. 2014; Soh, et al. 2014; Tomaszkiewicz, et al. 2016; Mahajan, et al. 2018; Hughes, et al. 2020). 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 & 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

79 techniques, like ATAC-seq (Buenrostro, et al. 2013), utilize short-read sequencing to profile accessible
80 chromatin regions (ACRs) at a fine-scale across the genome. Across autosomes, ACRs show signatures of
81 purifying selection relative to other intergenic regions, consistent with these regions containing important
82 functional elements for gene regulation (Connelly, et al. 2014; Lu, et al. 2019; Horvath, et al. 2021).
83 Profiling ACRs on sex chromosomes would therefore provide a means to study the molecular evolution
84 of *cis*-regulatory regions in the context of Y degeneration.

85 The threespine stickleback fish (*Gasterosteus aculeatus*) is an excellent model to study the
86 evolution of *cis*-regulatory regions on sex chromosomes. Threespine stickleback fish have a high-quality
87 reference assembly of the Y chromosome and this chromosome is more recently derived compared to
88 many other species with available chromosome-scale assemblies (Hughes, et al. 2010; Hughes, et al.
89 2012; Bellott, et al. 2014; Bellott, et al. 2017; Hughes, et al. 2020; Peichel, et al. 2020). The threespine
90 stickleback sex chromosomes contain multiple evolutionary strata at different stages of degeneration
91 (Roesti, et al. 2013; White, et al. 2015; Peichel, et al. 2020), allowing for comparisons of sequence
92 evolution at different temporal scales. Strata form if recombination is suppressed between sex
93 chromosomes in multiple steps (Bachtrog 2013). Each subsequent phase of recombination suppression is
94 followed by the accumulation of substitutions on the Y chromosome, leading to increased sequence
95 divergence between the X and Y. The age of individual strata can therefore be distinguished by varied
96 levels of synonymous divergence. In threespine stickleback fish, the oldest stratum is estimated to have
97 formed around 22 million years ago. Two younger strata formed between 4 and 6 million years ago
98 (Peichel, et al. 2020; Sardell, et al. 2021). Stratum one has undergone the most extensive degeneration,
99 with only 18% of ancestral coding regions remaining on the Y chromosome (Peichel, et al. 2020). Many
100 of the coding regions of these gametologs that remain in stratum one on the Y chromosome show
101 signatures of purifying selection and are enriched for dosage sensitive functions (White, et al. 2015;
102 Peichel, et al. 2020). Interestingly, chromosome-wide dosage compensation has not evolved to counter
103 the loss of expression for genes where the coding sequence has degenerated on the Y chromosome
104 (White, et al. 2015). The lack of chromosome-wide dosage compensation enables a gene-by-gene
105 comparison of *cis*-regulatory changes to understand how expression of X-linked alleles evolves in
106 response to degenerating coding regions on the Y.

107 Here, we leveraged ATAC-seq from two different tissues to identify *cis*-regulatory regions on the
108 sex chromosomes and an autosome. We found sex-linked *cis*-regulatory regions exhibited signatures of
109 positive selection, relative to non-functional, intergenic control regions. Our findings complement
110 existing sex chromosome evolution theory that suggests the accumulation of mutations within *cis*-
111 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 represent regions of the genome that are likely experiencing *cis*-regulatory evolution. We first sought to compare 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 three separate events, forming strata with distinct levels of divergence (stratum one: oldest; stratum two: intermediate; stratum three: youngest) (Peichel, et al. 2020). To survey ACRs, we utilized an Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) from liver tissue (Naftaly, et al. 2021) and testis tissue. We defined a set of non-coding 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 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 *(Pungititus 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 three 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 non-synonymous substitutions within coding regions across the sex chromosomes. Coding regions across all three strata on the threespine stickleback Y chromosome are accumulating non-synonymous substitutions at a higher rate than autosomes due to a reduced efficacy of selection (White, et al. 2015; Peichel, et al. 2020). We first compared the divergence among testis and liver ACRs to non-synonymous 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

145 non-synonymous mutations among all three strata (Figure 1) ($P < 0.001$; Kruskal-Wallis and Dunn's
146 Test). We also compared ACR divergence with synonymous substitutions, which is used as measure of
147 neutral evolution. Testis ACRs had divergence higher than neutral, synonymous substitutions in all three
148 strata (Figure 1) ($P < 0.001$; Kruskal-Wallis and Dunn's Test). Liver ACR's exhibited divergence that
149 was higher than synonymous substitutions in the two oldest strata (one and two) ($P < 0.001$; Kruskal-
150 Wallis and Dunn's Test) (Figure 1), indicating these regions are also evolving rapidly. In the youngest
151 stratum (three), the liver ACRs were not significantly different than synonymous substitutions ($P = 0.413$,
152 Kruskal-Wallis and Dunn's Test). Additionally, ACR divergence was distinct between all three
153 evolutionary strata ($P < 0.001$; Kruskal-Wallis and Dunn's test) and corresponded with the extent of
154 evolutionary divergence of each stratum. ACR divergence higher than both non-synonymous and
155 synonymous substitutions strongly suggests *cis*-regulatory regions are evolving rapidly across all three
156 strata. However, this analysis alone cannot rule out the possibility that purifying selection may still be
157 operating within coding regions on the Y chromosome, reducing non-synonymous substitutions.
158 Synonymous substitutions may also be under weak purifying selection (i.e. codon bias) (Bachtrog 2005;
159 Lawrie, et al. 2013). In this scenario, we may observe higher divergence rates within ACRs relative to
160 coding regions.

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

174 ACRs functional in the liver exhibited higher substitution rates on the Y chromosome than on the
175 X chromosome. Y-linked ACRs had significantly higher substitution rates compared to the non-functional
176 control regions for the two oldest evolutionary strata (Figure 2B; $P < 0.001$; stratum one and two; 10,000
177 permutations). In contrast to the Y-linked ACRs, we found the X-linked liver ACR sequence divergence

178 was significantly lower than non-functional control regions (Figure 2B; Stratum one: P = 0.016; Stratum
179 two and three: P < 0.001; 10,000 permutations). This suggests that ACRs may be functionally constrained
180 more on the X chromosome, compared to Y-linked ACRs. We observed a similar pattern on the Y
181 chromosome for ACRs functional in testis tissue. Y-linked ACRs had significantly higher substitution
182 rates in all three evolutionary strata, compared to non-functional control regions (Figure 2B; P < 0.001,
183 10,000 permutations). However, we observed X-linked ACRs from testis tissue, also exhibited elevated
184 nucleotide substitutions in the oldest two evolutionary strata (Figure 2B; P < 0.001; 10,000 permutations),
185 while the youngest stratum was significantly lower than intergenic regions (Figure 2B; P < 0.001; 10,000
186 permutations). This indicates some *cis*-regulatory elements exhibited accelerated substitution rates on the
187 X chromosome and this occurs in a tissue-specific fashion.

188 Autosomal ACRs have been shown to be under purifying selection in other species (Horvath, et
189 al. 2021). We therefore tested whether threespine stickleback autosomal ACRs also exhibited purifying
190 selection, similar to the pattern we observed on the X chromosome in liver. Consistent with patterns of
191 purifying selection, we found autosomal ACRs exhibited lower sequence divergence compared to non-
192 functional control regions in both liver (Figure 2B; P = 0.002; 10,000 permutations) and testis tissue
193 (Figure 2B; P < 0.001; 10,000 permutations).

194 **Accessible chromatin regions exhibit signatures of selection**

195 An excess of substitutions on the Y chromosome could be caused by relaxed purifying selection
196 or positive selection. To distinguish between these alternatives, we used short-read sequencing of 12
197 males from a freshwater population of stickleback fish (Shanfelter, et al. 2019) to search for signatures of
198 selection within ACRs. In the absence of selection, the ratio of polymorphism to sequence divergence
199 should be equal between functional and non-functional sites. Deviation from this expectation would imply
200 these regions are under selection. To explore this, we performed a modified McDonald-Kreitman (MK)
201 test on non-coding sequences (Andolfatto 2005). We compared the number of X- and Y- specific
202 substitutions to the number of polymorphisms segregating within a population for ACRs and non-
203 functional control sequences. Except for stratum two on the X chromosome, sex chromosome ACRs from
204 the liver and the testis had a greater proportion of sites fixed by positive selection ($\alpha$), compared to
205 autosomal ACRs (P < 0.001; chi-square test) (Table 1). We also identified the direction of selection
206 acting on ACRs by calculating the odds-ratio of the McDonald Kreitman test, known as the neutrality
207 index (NI) (Rand and Kann 1996). Liver ACRs of the sex chromosomes had an NI < 1 (P < 0.001; chi-
208 square test), consistent with positive selection (Figure 3; Table 1). The only exception we found was for
209 one of the younger strata on the X chromosome (stratum two), which had an elevated NI consistent with
210 purifying selection. Despite having a mean divergence lower than intergenic sequence, we found

211 signatures of positive selection on the X chromosome for strata one and three. In these cases, the
212 departure from neutrality was driven by the low number of polymorphisms within ACRs compared to
213 intergenic sequence, rather than high divergence within ACRs (Table 1). In these cases, the reduced
214 polymorphism may be due to recent positive selection (Stoletzki and Eyre-Walker 2010, Hughes, et al.
215 2008). Each stratum on the Y chromosome had a clear signature of positive selection (NI < 1; P < 0.001).
216 Compared to the X chromosome and the autosome, the Y chromosome liver ACRs had stronger
217 signatures of positive selection, indicative by a non-overlapping bootstrap interval. For the testis ACRs,
218 NI was indistinguishable between the X and Y chromosome, and both sex chromosomes had non-
219 overlapping bootstrap intervals with the representative autosome, indicating stronger positive selection
220 acting on sex-linked ACRs. Similar to X-linked ACRs in liver, we found that testis ACRs in stratum three
221 on the X chromosome exhibited an NI < 1, due to a low number of polymorphisms in ACR rather than an
222 excess in divergence.

223 We compared the patterns we observed across ACRs with coding regions. We found that coding
224 regions across the entire Y chromosome had an excess of non-synonymous polymorphisms (Figure 3; NI
225 > 1; P < 0.001; chi-square test), consistent with relaxed purifying selection on a degenerating Y
226 chromosome. This pattern was largely driven by the youngest two strata of the Y chromosome (Table 2).
227 The oldest stratum (one) did not deviate from neutral expectations (P = 0.112). Coding regions across the
228 X chromosome were not significantly different than neutral expectations (NI = 1; P = 0.281; chi-square
229 test). When split by evolutionary strata only the youngest stratum (three) deviated from neutral
230 expectations after correcting for multiple comparisons. In stratum three, the coding regions exhibited
231 purifying selection (NI > 1; Pn < Ps; P < 0.001; chi-square test). Unlike the Y chromosome, the X entire
232 chromosome did not have an excess of non-synonymous polymorphism (Table 2). A gene-by-gene MK
233 test revealed that 50% of X-linked gametologs had an NI > 1, compared to only 25% that had an excess of
234 amino acid substitutions (Supplemental material). Importantly, sex-linked coding regions were not under
235 widespread positive selection, like ACRs.

236 **Reduced polymorphism throughout the sex chromosomes**

237 Due to a lower effective population size, sex chromosomes are expected to have lower nucleotide
238 diversity relative to the autosomes, given an equal sex ratio in the population. To test whether nucleotide
239 diversity is reduced on the threespine stickleback sex chromosomes, we estimated average nucleotide
240 diversity in sliding windows across the sex chromosomes and autosome 18. Across windows, we found Y
241 chromosome nucleotide diversity was much lower than the expected 25% of nucleotide diversity
242 compared to autosomes (Y chromosome pi: 0.0001; autosome pi: 0.0043; Supplemental Figure 1). This
243 suggests that selection acting on linked deleterious or beneficial mutations has lowered nucleotide

244 diversity beyond neutral expectations on the Y chromosome. The X chromosome is predicted to contain
245 75% of the nucleotide diversity compared to the autosomes. We found that the non-PAR regions of the X
246 chromosome had 45% of the amount of nucleotide diversity as autosomes, indicating the X chromosome
247 may also be under selection, reducing nucleotide diversity beyond neutral expectations (non-PAR X
248 chromosome pi: 0.0019; autosome pi: 0.0043; Supplemental Figure 1). We found that the Y chromosome
249 had consistent signatures of low nucleotide diversity across all strata (Supplemental Figure 1). However,
250 there were distinct differences in nucleotide diversity for each stratum and the PAR on the X chromosome
251 (Supplemental Figure 1; $P < 0.001$ for all comparisons; Kruskal-Wallis and Dunn's Test), suggesting each
252 region of the X chromosome may be under different selection pressures depending on gene and regulatory
253 content.

254 **Y-linked substitutions are correlated with biased expression from the X chromosome**

255 Selection should favor the loss of expression from the Y chromosome in order to silence coding
256 regions that have accumulated deleterious mutations (Orr and Kim 1998; Bachtrog 2006; Lenormand, et
257 al. 2020). To test if Y-linked *cis*-regulatory divergence was associated with changes in expression, we
258 compared ACR sequence divergence to changes in allelic expression on the X and Y from liver and testis
259 RNA-seq transcriptomes (Peichel, et al. 2020). To maximize the total number of genes for this analysis,
260 we pooled genes across all three evolutionary strata. We found that sequence divergence among liver
261 ACRs that were proximal to genes was a predictor of allele-specific expression. We observed higher
262 expression of the X-linked gametolog, relative to the Y-linked gametolog, when ACRs on the Y
263 chromosome had more substitutions (Figure 4A; N = 51; P = 0.004; R=0.4; Spearman's rank correlation;
264 95% Confidence Interval via bootstrapping (0.1249, 0.5769)), indicating that ACR divergence explains
265 around 16% of the X-biased expression observed in liver tissue We also observed a correlation between
266 expression and substitutions within ACRs on the X chromosome (Figure 4C; N = 51; P = 0.033; R=0.30;
267 Spearman's rank correlation; 95% Confidence Interval from bootstrapping (0.0909, 0.4742)). The slopes
268 of each linear regression were indistinguishable between the X and Y chromosome (P =0.590; Fisher Z
269 transformation). Although we do not have gene expression from the ninespine stickleback to determine
270 which sex chromosome is deviating from the ancestral expression level, the Y-linked substitutions
271 indicate the X-biased expression we observe is likely due to downregulating the Y chromosome.

272 A similar trend was observed in testis tissue for Y-linked ACR divergence (Figure 4), but the
273 correlations were not significant (Figure 4B; N = 31; P = 0.26; R=0.201; Spearman's rank correlation;
274 95% Confidence Interval via bootstrapping (-0.0095, 0.5706)). There are fewer testis-expressed genes in
275 this analysis. Therefore, the lack of significance could be due to sample size. To identify if sample size
276 had an effect, we performed a randomized down-sampling for the liver-expressed genes. Even with down-

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

282 ***Cis*-regulatory evolution is not correlated with deleterious mutations in coding regions**

283 The downregulation of Y-linked gametologs could be adaptive if loss of expression follows the
284 accumulation of deleterious coding substitutions (Orr and Kim 1998). We searched for evidence of
285 adaptive silencing on the Y chromosome by comparing the ACR divergence to the ratio of non-
286 synonymous to synonymous substitutions ($d_N/d_S$) within the coding region of each gene. An adaptive
287 silencing model would be supported if Y-linked gametologs with an elevated $d_N/d_S$ ratio also have
288 elevated *cis*-regulatory divergence. Among genes shared between the X and Y chromosomes, we found
289 no correlation between ACR nucleotide divergence and $d_N/d_S$ (Supplemental Figure 2; each stratum on X
290 and Y: $P > 0.05$; Spearman's rank correlation), indicating coding regions with elevated ACR divergence
291 are not more likely to have elevated substitutions in the amino acid sequence.

292 Elevation of $d_N/d_S$ on the Y chromosome could indicate relaxed purifying selection, or positive
293 selection. Combining genes that are evolving under both could mask signatures of adaptive silencing. To
294 account for this, we split gametologs on the Y chromosome into classes based on the signatures of
295 selection from the MK test. We grouped gametologs into categories of purifying selection ($NI_{Haldane} >$
296 $0.25$; $N = 262$), positive selection ($NI_{Haldane} < -0.25$; $N = 41$), and genes evolving through relaxed
297 purifying selection ($-0.25 > NI_{Haldane} > 0.25$; $N = 219$). If selection was acting on ACRs to downregulate
298 genes with deleterious mutations, we would expect to see a positive correlation between ACR divergence
299 and $d_N/d_S$ for genes with signatures of relaxed purifying selection. In contrast to this prediction, we found
300 a weak negative correlation between ACR divergence and $d_N/d_S$ for expressed genes under relaxed
301 purifying selection (Supplemental Figure 3; $R= -0.293$; $P =0.0155$; $N=35$; Spearman's rank correlation).
302 These genes also did not have more X-biased expression than other genes (Supplemental Figure 4; $P >$
303 $0.05$; Kruskal-Wallis test). Interestingly, we did find a positive correlation between ACR divergence and
304 $d_N/d_S$ for genes evolving under positive selection (Supplemental Figure 4; $R=0.92$; $P=0.002$; $N=10$;
305 Spearman's rank correlation). Overall, most genes on the Y seem to be evolving independently from
306 neighboring regulatory regions, except for a small set of genes under positive selection.

307 We also examined whether ACR divergence was higher within genes that contained frameshift
308 mutations, nonsense mutations, or in-frame deletions. These types of mutations are more likely to produce

309 a non-functional or sub-optimal peptide. We found no significant difference in the number of ACR
310 substitutions from liver and testis tissue for coding regions that contained frameshift, nonsense mutations,
311 or deletions, compared to coding regions that did not have these putatively deleterious mutations
312 (functional) (Supplemental Figure 5) (all pairwise comparisons $P > 0.05$; Mann-Whitney U test). Overall,
313 we found no evidence of *cis*-regulatory divergence associated with deleterious coding sequence
314 divergence, suggesting the adaptive silencing model is unlikely the major driver of *cis*-regulatory
315 evolution on the threespine stickleback sex chromosomes.

316

317 **Discussion**

318 The evolution of gene regulation on Y chromosomes has largely been studied in the context of
319 RNA expression levels of gametologs (Muyle, et al. 2012; White, et al. 2015; Beaudry, et al. 2017;
320 Muyle, et al. 2018; Martin, et al. 2019; Veltsos, et al. 2019). Over time, gametolog expression is generally
321 lost across most genes on the Y chromosome. With the growing number of Y chromosome reference
322 assembles and sequencing methods to quickly profile functional *cis*-regulatory regions across the genome,
323 it has become feasible to explore the molecular evolution of *cis*-regulatory regions that have led to altered
324 expression patterns. Here, we observed elevated nucleotide substitutions and reduced polymorphisms
325 within *cis*-regulatory elements across the threespine stickleback Y chromosome, consistent with positive
326 selection driving this process. We found that increased rates of *cis*-regulatory divergence were associated
327 with X-biased expression of gametologs, suggesting loss of expression from the Y chromosome. Without
328 liver and testis gene expression from the ninespine stickleback as an ancestral comparison, we were
329 unable to confirm that X-biased gametolog expression is a result of downregulation of the Y chromosome
330 or upregulation of the X chromosome. However, previous sequencing of a brain transcriptome showed
331 expression patterns consistent with loss of Y expression rather than gain of X expression (White, et al.
332 2015). It is therefore likely that the X-biased expression we observed in liver and testis tissue also occurs
333 through the loss of Y chromosome expression, similar to what has been observed in other species with
334 degenerating Y chromosomes (Muyle, et al. 2012; Beaudry, et al. 2017; Wei and Bachtrog 2019).
335 Additional work will be necessary to identify the functional variants responsible for the X-biased
336 expression pattern. We currently cannot determine whether the X-biased expression pattern is due to the
337 accumulation of deleterious mutations from inefficient selection, adaptive substitutions that are
338 accumulating in the ACRs, or a combination of the two.

339 Several theoretical models of sex chromosome evolution have been developed that predict how
340 gene expression evolves over time. Like coding regions, *cis*-regulatory regions will accumulate

341 deleterious mutations because of a reduced efficacy of purifying selection (Charlesworth and
342 Charlesworth 2000; Bachtrog 2008). Without the removal of deleterious mutations, substitution rates
343 within regulatory and coding regions approach the rates observed among non-functional intergenic sites.
344 Theory predicts that substitution rates within regulatory regions can be elevated further through positive
345 selection favoring beneficial mutations. The adaptive silencing model (Orr and Kim 1998) posits that Y
346 gametologs that have accumulated maladaptive coding substitutions would be favored to be
347 downregulated through *cis*-regulatory mutations. Only the functional X-linked gametolog would be
348 expressed in males. An alternative model suggests that deleterious mutations accumulate in *cis*-regulatory
349 elements throughout the non-recombining region, initially downregulating all Y-linked gametologs (i.e.
350 degeneration by regulatory evolution (Lenormand, et al. 2020; Lenormand and Roze 2022). As
351 deleterious mutations simultaneously accumulate within coding regions, a positive-feedback loop ensues,
352 leading to a higher substitution rate within *cis*-regulatory elements to selectively downregulate
353 increasingly maladaptive coding regions. Our survey of ACR evolution in threespine stickleback fish
354 allowed us to compare with each of the previously discussed theories.

355 Evidence for adaptive silencing (Orr and Kim 1998) would be present if alleles with elevated
356 rates of coding divergence between the X and Y chromosomes, were more likely silenced on the Y
357 chromosome. Our results did not support this model. We found that the accumulation of X or Y ACR
358 divergence had no association with relaxed purifying selection in nearby coding regions. Our results
359 complement previous findings that have compared coding sequence evolution to gametolog expression.
360 The level of expression from the Y chromosome is often not correlated with the overall number of
361 deleterious mutations within coding regions (Bachtrog 2006; Bachtrog, et al. 2008; Beaudry, et al. 2017).
362 While an elevation of $d_N/d_S$ has been previously used as a proxy for the accumulation of deleterious
363 variants on the Y, this association could also be conflated by positive selection. Here, we showed the lack
364 of an association still held when only looking at genes evolving under relaxed selection. Additionally, we
365 found no correlation between divergence within an ACR and the functional state of a coding region (i.e.
366 containing frameshift or nonsense mutations). We found *cis*-regulatory evolution was correlated with
367 expression differences, but that ACR divergence accumulated independent of coding divergence. A
368 correlation between nucleotide substitutions in *cis*-regulatory regions and deleterious mutations in coding
369 regions may be observed if ACRs and gene expression are profiled from additional tissues. In addition,
370 we assigned ACRs to coding sequences based on overall proximity to a gene. Although this method of
371 annotating ACRs is commonly used to assign *cis*-regulatory regions to genes (Connelly, et al. 2014;
372 Alexandre, et al. 2018; Ricci, et al. 2019), it is possible that some ACRs interact with other genes through
373 long range interactions based on chromatin configuration (Mifsud, et al. 2015; Schoenfelder and Fraser
374 2019), reducing our ability to detect an association between deleterious mutations in coding regions and

375 nucleotide substitutions in ACRs. Also, the number of substitutions within an ACR may not be a fully
376 accurate predictor of the overall ability to silence a given gametolog on the Y chromosome. If only a
377 small number of substitutions are needed to ablate transcription factor binding within an ACR, we would
378 not expect to see a strong correlation between the number of deleterious mutations within coding
379 sequences and the number of substitutions within ACRs. Additional functional work will be necessary to
380 explore these alternatives.

381 Elevated substitution rates in *cis*-regulatory regions quickly following recombination suppression
382 would be consistent with the model of degeneration by regulatory evolution (Lenormand, et al. 2020;
383 Lenormand and Roze 2022). Silencing of Y-linked alleles occurs first in this model, followed by the
384 accumulation of deleterious mutations within coding regions. We found that putative *cis*-regulatory
385 elements exhibited signatures of positive selection on the Y chromosome across all three evolutionary
386 strata, including the two youngest strata that still contain most of the ancestral gene content (Peichel, et al.
387 2020). The degeneration by regulatory evolution model predicts that the substitution rate within *cis*-
388 regulatory elements will accelerate once deleterious mutations begin to accumulate within coding regions.
389 This will further suppress gametolog expression from the Y chromosome, masking maladaptive coding
390 mutations. Consistent with this model, our results show *cis*-regulatory regions have stronger signatures of
391 positive selection compared to autosomal ACRs, as well as both coding and non-coding regions on the
392 sex chromosomes. This suggests there is selection to downregulate Y-linked alleles in liver. Alternatively,
393 substitutions in ACRs could reflect adaptation to upregulate genes on the Y chromosome, to maintain
394 expression or to gain male beneficial expression (Skaletsky, et al. 2003; Martínez-Pacheco, et al. 2020;
395 Shaw and White 2022). Non-adaptive explanations should also be carefully considered as alternatives to
396 adaptive models. For example, elevated substitution rates could be observed in Y-linked ACRs if they
397 have higher mutation rates compared to synonymous and silent intergenic sites. Indeed, it has been
398 suggested that male-biased mutation rates, due to increased cell division (Link, et al. 2017), or specialized
399 DNA repair pathways (Chang, et al. 2022), could shape the evolution of the Y chromosome. These
400 mutations could fix at a faster rate within populations and produce false adaptive signatures because of
401 the smaller effective population size of the Y chromosome. However, the threespine stickleback fish Y
402 chromosome does not show evidence of male-biased mutation rate at synonymous sites (White, et al.
403 2015), and thus it remains unclear how these mutation biases would increase substitution rates uniquely at
404 accessible chromatin regions.

405 The degeneration by regulatory evolution model also indicates dosage compensation should
406 evolve simultaneously for genes that are under strong stabilizing selection to maintain expression levels
407 (Lenormand, et al. 2020). Degeneration of *cis*-regulatory elements on the Y chromosome and loss of

408 expression should select for upregulation of the gametolog on the X chromosome to compensate for
409 dosage loss. In this scenario, nucleotide substitutions should accumulate within ACRs on the X
410 chromosome. We found that X-linked ACRs had a depletion of polymorphism, consistent with positive
411 selection. Unlike Y-linked ACRs, the X-linked ACRs had low divergence over longer evolutionary
412 timespans between ninespine and threespine stickleback fish. This suggests that selection has acted on
413 some X-linked sites more recently within threespine stickleback populations to fix beneficial regulatory
414 mutations. This distinction is consistent with Y-linked ACR divergence accumulating first, as predicted
415 by the degeneration by regulatory evolution model (Lenormand et al. 2020). Selection on X-linked *cis*-
416 regulatory elements within ACRs may still be ongoing. A similar finding was observed on the neo-sex
417 chromosomes of *Drosophila miranda*, where the neo-X chromosome was enriched for signatures of
418 selective sweeps, while the ancestral X chromosome region was not (Bachtrog, et al. 2009).

419 We found the strongest signatures of positive selection on the X chromosome were from testis
420 ACRs. Testis ACRs may have faster rates of sequence divergence on both the X and Y chromosome,
421 from processes similar to the faster X-effect (Charlesworth, et al. 1987; Vicoso and Charlesworth 2009).
422 There is widespread evidence for accelerated rates of coding evolution on X chromosomes compared to
423 autosomes, especially in male specific tissues (Baines, et al. 2008; Meisel and Connallon 2013; Parsch
424 and Ellegren 2013; Larson, et al. 2016; Kopania, et al. 2022), as the X-linked mutations in hemizygous
425 regions are always exposed to selection in males. Expression of X-linked gametologs may also evolve at
426 faster rates (Khaitovich, et al. 2005; Brawand, et al. 2011; Meisel, et al. 2012b; Coolon, et al. 2015;
427 Kopania, et al. 2022), presumably due to mutations in *cis*-regulatory regions. The elevated substitution
428 rate we observed on the X chromosome in testis ACRs may reflect male-beneficial selection to alter
429 expression levels that is not necessarily connected to dosage compensation.

430 While we observed molecular signatures of positive selection on ACRs, we did not find sufficient
431 evidence to suggest selection is associated with elevated X expression, as predicted by the degeneration
432 by regulatory evolution model. While selection for dosage compensation could be ongoing, previous
433 findings suggest that most genes on the threespine stickleback Y chromosome can be lost without
434 consequence. Within the oldest stratum, over half of the Y-linked gametologs have been completely lost
435 and males exhibit half expression, relative to females (White, et al. 2015). Even across species with
436 chromosome-wide dosage compensation mechanisms, many genes are not compensated completely
437 (Cotton, et al. 2013; Tukiainen, et al. 2017), which suggests that loss of Y expression is not always
438 deleterious. The dosage sensitivity of each sex-linked gene may be essential for predicting how rapidly
439 *cis*-regulatory regions evolve on both sex chromosomes. In the degeneration by regulatory evolution
440 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.

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; Bellott, et al. 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 with 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 three 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 upregulation. 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 sub-functionalization (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*-

474 regulatory regions associated with testis-specific genes. Interestingly, we found signatures of positive
475 selection in ACRs near testis-expressed genes. Some of these genes were enriched for Y-biased
476 expression. These findings provide evidence that ancestral gametolog expression levels can also change
477 presumably through gain of new testis-related functions. Additional work will be necessary to identify the
478 novel function these regulatory regions provide during spermatogenesis. An increase in testis expression
479 was also found for ancestral gametologs in a comparative study across 17 species of mammals (Martínez-
480 Pacheco, et al. 2020). In these species, most Y-linked gametologs gained novel testis expression patterns,
481 compared to their X-linked alleles. Rapid sex-linked *cis*-regulatory evolution may therefore be a universal
482 phenomenon across species.

483 Y chromosomes often have much lower diversity overall, relative to neutral expectations
484 (Lawson Handley, et al. 2006; Wilson Sayres, et al. 2014). This pattern has been attributed to purifying
485 selection removing deleterious mutations and the linked neutral variation throughout the non-recombining
486 region (Wilson Sayres, et al. 2014; Wilson Sayres 2018). Our findings revealed that nucleotide diversity
487 on the threespine stickleback Y chromosome is also much lower than neutral expectations. Our results
488 suggest positive selection within ACRs could also be an important driver of reducing nucleotide diversity
489 on the Y chromosome within populations. One important consideration is whether sex-biased
490 demography may affect Y-linked diversity. If males are more variable in reproductive success than
491 females, this could lower the expected effective population size of the Y chromosome, reducing
492 nucleotide diversity. Little is known about the operating sex-ratios of stickleback fish. Some populations
493 exhibit sex ratios that are female biased (Rollins, et al. 2017). This could contribute to the low nucleotide
494 diversity we observed. However, simulations in human populations revealed that even drastic shifts in sex
495 ratio could not entirely explain the low diversity observed on the Y chromosome (Wilson Sayres, et al.
496 2014).

497

498 **Conclusion**

499 Together, our results provide evidence of positive selection driving accelerated rates of nucleotide
500 substitution in *cis*-regulatory elements. Signatures of positive selection, even in the youngest two strata,
501 indicate that *cis*-regulatory evolution can proceed rapidly following the suppression of recombination,
502 leading to reduced gene expression from the Y chromosome. These results support some aspects of recent
503 models of sex chromosome evolution, where *cis*-regulatory degeneration and silencing occurs first on the
504 Y chromosome. However, we found this can occur in the absence of dosage compensation, contrasting
505 other assumptions of these models. Improvements in functional annotations of regulatory regions as well

506 as an ever-growing collection of high-quality Y and W assemblies will allow continued empirical testing
507 of new regulatory models of sex chromosome evolution.

508

509 **Materials and Methods**

510 **Molecular evolution of accessible chromatin regions and coding regions**

511 We used accessible chromatin regions (ACRs) from two different tissues. Liver ACRs were
512 previously identified from two replicates (Naftaly, et al. 2021) (NCBI BioProject PRJNA667175). We
513 also collected testes from two juvenile males (~4.4 cm in standard length) of laboratory-reared threespine
514 stickleback fish, originally isolated from Lake Washington (Seattle, Washington, USA) (NCBI BioProject
515 PRJNA686097). The testis cells were immediately dissociated through homogenization in 1X PBS
516 containing proteinase inhibitor cocktail (PIC, cOmplete tablets Roche) (PBS+PIC). The cells were fixed
517 with 16% formaldehyde and washed twice with PBS+PIC, followed by lysis in 1M Tris-HCl, pH=8, 0.5M
518 EDTA, 10% NP-40, 50% glycerol/molecular grade $H_2O$, and 1X PIC. Nuclei were stained with DAPI and
519 counted with a hemocytometer. We diluted samples to 60,000 – 80,000 nuclei. ATAC-seq library
520 preparation was conducted using previously established protocols (Lu, et al. 2017; Naftaly, et al. 2021).
521 The libraries were sequenced on an Illumina NextSeq (2 x 150 bp; Georgia Genomics and Bioinformatics
522 Core). Reads were trimmed with Trimmomatic (v. 0.36) (Bolger, et al. 2014), using a sliding window of
523 four bases, trimming the remainder of the read when the average quality within a window dropped below
524 20. Residual adapter sequences were removed using Trimmomatic ILLUMINACLIP. Reads were filtered
525 for a minimum length using MINLEN:30. We aligned the trimmed reads using Bowtie2 (v. 2.4.1) with
526 default parameters (Langmead and Salzberg 2012). We filtered for alignments with a mapping quality
527 greater than 20 using SAMtools (v1.14). We also removed PCR duplicates by using the MarkDuplicates
528 function in Picard (https://github.com/broadinstitute/picard).

529 We used a Tn5 control sample to normalize ATAC-seq reads to remove the effect of Tn5 bias.
530 For the control, genomic DNA was extracted from a caudal fin clip of one male fish using a standard
531 phenol-chloroform extraction. The ATAC-seq library preparation, whole genome sequencing library
532 preparation, and sequencing (Illumina HiSeq 2 x 150 bp) of the control sample were completed by
533 GENEWIZ (New Jersey, USA). Whole genome sequencing was conducted in order to show the DNA
534 sample was of sufficient quality to construct the Tn5 bias control. For the whole genome sequencing, we
535 recovered over 232 million reads with an average quality score of 38.59, indicating a high-quality sample.
536 We trimmed residual adapters and low-quality sequences using Trimmomatic as previously described.
537 Trimmed reads were aligned to the threespine stickleback genome using Bowtie2. The read coverage per

538 base pair was calculated using BEDTools (v2.29, -d) (Quinlan and Hall 2010). The whole genome
539 sequencing sample had an average read depth of 90x where only ~10% of the genome was supported with
540 less than ten reads. Within these regions, only 6% had zero reads per base pair, indicating 94% of the
541 genome could be queried for biased integration of Tn5. The Tn5 bias control produced over 177 million
542 reads. The reads were trimmed from these reads using Trimmomatic with the same parameters. The
543 trimmed reads were aligned to the threespine stickleback genome using Bowtie2 with default parameters.
544 In order to use previous alignment coordinates between the X and Y chromosome, we used the v. 4
545 genome build of the X chromosome (Peichel, et al. 2020). We used the v. 5 genome build of chromosome
546 18 (Nath, et al. 2021). Reads mapping to the mitochondria and unscaffolded regions were removed. PCR
547 duplicates were removed using MarkDuplicates from Picard.

548 We evaluated concordance between replicates in several ways. We first searched for enrichment
549 of ATAC-seq reads around the transcription start sites (TSSs) of expressed genes. Iso-Seq long read
550 sequencing was previously conducted in threespine stickleback fish to curate accurate transcription start
551 site annotations across multiple tissues (Naftaly, et al. 2021). We used deepTools (v3.5.1) computeMatrix
552 reference-point (Ramírez, et al. 2016) on the ATAC-seq alignments to assay read depth 3kb around the
553 complete set of annotated transcription start sites. We plotted enrichment of ATAC-seq reads around TSS
554 and compared to RNA-seq expression using deepTools plotHeatmap with default settings, except that we
555 used the -sortUsingsamples to sort the regions by expression. We found strong enrichment around TSSs,
556 across both replicates for the same set of expressed genes (Supplemental Figure 6). ACRs were called for
557 each replicate using MACS2 (v2.2.7.1) callpeak with the –keep--dup all parameter, and read depth was
558 normalized with Tn5 control sequencing with the -c parameter (Zhang, et al. 2008). The majority of
559 ACRs were shared across replicates (liver: 67% and testis: 61%) We then used the P-value of each
560 MACS2 peak to calculate the Irreproducibility Discovery Rate (IDR) (https://github.com/daniel-
561 shaw1/Regulatory_divergence_paper/blob/main/Supplemental/IDR.r), (Supplemental Figure 7). The IDR
562 determines the probability that a peak is reproducible between replicates, due to chance (Qunhua, et al.
563 2011; Landt, et al. 2012). Replicates that have IDR < 0.05 are generally considered to be highly
564 reproducible. However, the IDR test is considered very stringent (Jalili, et al. 2015), potentially removing
565 many true regions of increased accessibility. Because both replicates were exhibited high concordance,
566 we pooled the aligned reads from both replicates and called peaks from the joint alignment using MACS2
567 as previously described, as pooling raw reads has been previously reported (Reske, et al. 2020; Yan, et al.
568 2020) as a strategy to find high quality enriched regions. We found that most ACRs (liver: 79%, testis:
569 75%) from the pooled sample overlapped with low scoring ACRs from the individual replicates (IDR <
570 0.05). This suggested a majority of ACRs in the pooled dataset were of high quality in the individual
571 replicates.

572 To define an ancestral set of ACRs, we mapped the nucleotide sequence of each non-coding ACR
573 on the X chromosome to the homologous region on the Y chromosome using previously generated
574 alignment coordinates (Peichel, et al. 2020). A similar approach has been used previously to define the
575 ancestral set of coding regions, aligning X-linked gene annotations to the Y chromosome (Peichel, et al.
576 2020). We also attempted to identify Y-linked ACRs and calculate divergence in a similar method as we
577 identified with X-linked ACRs. However, we found that MACS2 identified significantly fewer enriched
578 regions of accessible chromatin on the Y chromosome compared to the X chromosome. The few ACRs
579 that were identified also aligned to multiple regions on the X chromosome, suggesting these regions were
580 repetitive. Using the X chromosome ACRs, we identified orthologous regions between the X
581 chromosome, Y chromosome, and orthologous autosome (chromosome 19) from the ninespine
582 stickleback fish *(Pungitius pungitius*; (Varadharajan, et al. 2019) using BLAST+ (blastn v. 2.11.0) with
583 default blastn parameters and -perc_identity set to 75. (Camacho, et al. 2009). To ensure high quality
584 alignments, we filtered for uniquely mapping alignments that also had a bit score greater than 100. We
585 used BLAST+ alignments to extract ACRs and the proximal 50bp upstream and downstream, to create
586 multiple sequence alignments for downstream analysis. We aligned the threespine stickleback X and Y
587 ACR sequences to the ninespine stickleback autosome sequence using MUSCLE (v 3.8.1551) (Edgar
588 2004) with default parameters. We used MUSCLE to create a multiple sequence alignment between the X
589 sequence, Y sequence, and the sequence from the orthologous autosome in order to identify X- and Y-
590 specific variants. We called single nucleotide variants using snp-sites (v2.5.1) (Page, et al. 2016) with
591 default parameters. We calculated the reported divergence rate by dividing the number of variants by the
592 number of the aligned sites.

593 We compared ACR divergence on the sex chromosomes to ACR divergence on autosome 18, a
594 chromosome similar in length to the X chromosome. We identified orthologous regions between the
595 threespine stickleback autosome 18 and autosome 18 from the ninespine stickleback fish using BLAST+
596 (blastn v. 2.11.0) (Camacho, et al. 2009) as previously described. We called variants between autosomal
597 regions in a similar manner as the sex chromosomes, by creating a pairwise alignment using MUSCLE (v
598 3.8.1551) with default parameters and called single nucleotide variants using the same python script as
599 above. We calculated a divergence rate by dividing the number of single nucleotide variants by the size of
600 the aligned region.

601 Divergence within ACRs was compared with the divergence of coding regions of neighboring
602 genes. We identified the closest gene to each ACR by running annotatepeaks.pl in homer (v 4.11) (Heinz,
603 et al. 2010) using default settings. Per default settings, ACRs were assigned to the closest gene within
604 50kb of the TSS. We verified the homer peak annotations with a custom Python script that identifies

605 ACRs near the TSS of genes (https://github.com/daniel-
606 shaw1/Regulatory_divergence_paper/blob/main/Supplemental/Peaksneargenes.py). Both methods were
607 100% concordant with each other. We further verified a subset of the annotations (200 peaks total) by
608 visualizing the ACRs in the threespine stickleback JBrowse genome browser
609 (https://stickleback.genetics.uga.edu/). We found that 100% of the 200 X-linked ACRs were located
610 proximal to the predicted annotated gene (Supplemental Figure 8). For coding divergence comparisons,
611 we used previously reported estimates of synonymous ($d_S$) and non-synonymous ($d_N$) divergence of
612 gametologs between the X and Y chromosomes (Peichel et al., 2020). Coding regions with $d_N/d_S$ ratios of
613 99 were omitted as these represent alignments with very little sequence divergence to estimate $d_N/d_S$
614 accurately.

615

616 **Estimating a control substitution rate**

617 To generate control regions, we randomly sampled non-functional intergenic regions throughout
618 the X chromosome, Y chromosome, or chromosome 18 to estimate a neutral substitution rate. Regulatory
619 regions tend to be GC-rich, which can be prone to higher mutation rates through mechanisms like GC-
620 biased gene conversion or spontaneous deamination of methylated cytosines (Nesta, et al. 2021). We
621 therefore GC-matched the randomly drawn intergenic regions with the GC content of ACRs. We
622 calculated the GC percentage of ACRs using a modified Perl script (countbp.pl, Nicholas Navin,
623 http://www.navinlab.com/bioperl/bioperl/gc_content.html). For each ACR, a random intergenic region
624 was drawn from the X chromosome equal in length and with a GC content within 2.0%. Intergenic
625 regions were defined as any region that fell outside of annotated functional regions from a combination of
626 Ensembl annotations (Cunningham, et al. 2021), Isoseq transcripts (Naftaly, et al. 2021), Y chromosome
627 annotations (Peichel, et al. 2020), repetitive elements (Peichel, et al. 2020; Nath, et al. 2021) and ACRs
628 from this study. This was performed for each stratum individually using bedtools shuffle –I ACRs.bed -
629 incl XYalignment.bed -excl annotations.bed. We extracted 10,000 sets of GC-matched intergenic regions
630 from each of the three evolutionary strata on the threespine stickleback sex chromosomes (set size same
631 as number of Liver ACRs: stratum one: 442, stratum two: 228, stratum three: 199; Testis ACRs: stratum
632 one: 30, stratum two: 37, stratum three: 51). Due to a limited number of intergenic sites shared between
633 the sex chromosomes, we sampled with replacement to align a sufficient number of intergenic regions.
634 We identified the orthologous regions to the ninespine genome assembly using BLAST+ (blastn v2.11.0)
635 (Camacho, et al. 2009). We filtered for alignments that mapped uniquely to chromosome 19 in the
636 ninespine stickleback fish, had an alignment length equal to 75% of the threespine query sequence, and
637 had a bit-score greater than 100. We generated multiple sequence alignments using MUSCLE (v

638 3.8.1551), for each intergenic permutation and calculated X and Y specific substitution rate as previously
639 described.

640

641 **Estimating within population nucleotide diversity**

642 We used whole-genome short-read sequencing of 12 males from the Lake Washington population
643 (Washington, USA; NCBI SRA SRP137809) to estimate nucleotide diversity across the sex
644 chromosomes. The raw reads were trimmed with Trimmomatic (v. 0.39) (Bolger, et al. 2014), using a
645 sliding window of four bases, trimming the remainder of the read when the average quality within a
646 window dropped below 15. The leading and trailing base pairs below quality three of every read were
647 removed along with any residual adapter sequence. After trimming, any reads below a minimum length of
648 36 were discarded. We aligned the trimmed reads using Bowtie2 (v. 2.4.5) with default parameters
649 (Langmead and Salzberg 2012). We also marked PCR duplicates by using the MarkDuplicates function in
650 Picard (v. 2.26.10) (https://github.com/broadinstitute/picard). SNP genotyping was conducted using the
651 GATK software package (v. 4.2.5.0). We called variants using HaplotypeCaller in GVCF mode. We then
652 joint-called variants using GenotypeGVCFs using a genomics database created by GenomicsDBImport.

653 We only considered biallelic SNPs in our estimates of nucleotide diversity and we filtered for
654 high quality genotypes using several different methods. On the sex chromosomes, sites should only have
655 hemizygous genotypes in males in the non-crossover region outside of the pseudoautosomal region. Any
656 sites that are heterozygous would be caused by errors in read alignment. We therefore did not consider
657 sites that were heterozygous in the non-crossover region on the X or Y chromosomes. We also filtered out
658 sites that exhibited too low or high of read depth, which would be indicative of alignment errors. To do
659 this, we did not consider sites that were less than one-half or greater than double the median read depth of
660 each chromosome (X and Y chromosomes: positions were retained if the read depth was between 3.5 and
661 14; chromosome 18: positions were retained if the read depth was between 6.5 and 26). Read filtering was
662 conducted using custom Perl scripts.

663 Nucleotide diversity (pi) was estimated on a per-site basis (--site-pi) across each chromosome
664 using vcftools (v. 0.1.16) (Danecek, et al. 2011)). Nucleotide diversity was averaged within 10,000 bp
665 non-overlapping windows across chromosomes X, Y, and 18.

666

667 **Testing for positive selection using population polymorphism**

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

684 We also performed a McDonald-Kreitman test on coding regions. We aligned the coding
685 sequence of all threespine stickleback X chromosome Ensembl predicted genes (build 95) to the Y
686 chromosome and the ninespine stickleback outgroup (autosome 19; v. 7) (Varadharajan, et al. 2019) using
687 Exonerate (v. 2.4.0) (Slater and Birney 2005) with parameters: --model coding2genome -M 2000 --bestn
688 1 --annotation. The Ensembl coding regions from the threespine stickleback autosome 18 were also
689 aligned to the ninespine stickleback autosome 18. The highest scoring alignment from Exonerate was
690 used for the MK test. Divergence and polymorphism was tallied at each position of the alignment. If a
691 position was variable, we determined if this was a synonymous or non-synonymous mutation in the
692 threespine stickleback by comparing to the ninespine stickleback translated codon. Sites were omitted if
693 they contained any N characters or contained any missing genotype data among the 12 males within the
694 VCF file. Codons were omitted if they contained any gaps from the alignment. We only considered
695 biallelic sites when counting polymorphisms and fixed substitutions within the Lake Washington
696 population. We used custom Perl scripts to count the total number of divergent and polymorphic sites. To
697 compare to ACRs, we calculated $\alpha$ and NI by pooling sites across all coding regions. For gene-by-gene
698 MK tests, we log transformed NI to include genes with a count of zero for any particular category of
699 divergence or polymorphism and refer to the value as $NI_{Haldane}$ (Haldane 1956; Stoletzki and Eyre-
700 Walker 2010).

701

702 **Comparison with allele-specific expression patterns**

703 We measured allele-specific expression of transcripts on the X and Y chromosome using three
704 replicate liver samples (NCBI BioProject PRJNA591630) and three replicate testis samples (NCBI
705 BioProject PRJNA591630). We trimmed the RNA-seq reads with Trimmomatic (v. 0.36) (Bolger, et al.
706 2014) using a sliding window of 4 bases, trimming the remainder of the read when the average quality
707 within a window dropped below 20 (SLIDINGWINDOW:4:20). Residual sequencing adapters were also
708 removed using ILLUMINACLIP. We aligned RNA-seq reads using Tophat2 (v2.2.1) with default
709 parameters (Kim, et al. 2013). We filtered for alignments with a map quality greater than 25 using
710 SAMtools (v1.14) (Li, et al. 2009) to identify reads that map uniquely on the X and Y allele. Similar map
711 quality filters have been used to distinguish short reads between sex chromosomes in *Drosophila* and
712 stickleback fish (Ellison and Bachtrog 2019; Peichel, et al. 2020) and between sexes with heteromorphic
713 sex chromosomes in guppies (Kirkpatrick, et al. 2021). Read counts were obtained using htseq-count (v
714 0.9.1) (Anders, et al. 2015). Default parameters were used with the addition of --stranded=no and --
715 nonunique all to maximize our ability to count reads on the Y chromosome. We used a custom GTF file
716 for htseq-count that included previously determined start and end sites of all Ensembl predicted
717 transcripts that are shared between the X and Y chromosomes (Peichel, et al. 2020). Reads were counted
718 on the X and Y across the entire transcript as a feature. Transcripts were removed from the analysis if
719 they had an RNA expression count of zero in all samples. For each tissue, we calculated the average read
720 count for each gametolog across the three tissue replicates. We calculated the allele-specific expression as
721 $\log_2$(X transcript average / Y transcript average).

722

723 **Data availability**

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

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

728

729 **Acknowledgements**

This work was supported by the National Science Foundation MCB 1943283 (M.A.W.), the National Institutes of Health R01GM147312 (M.A.W.), 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.

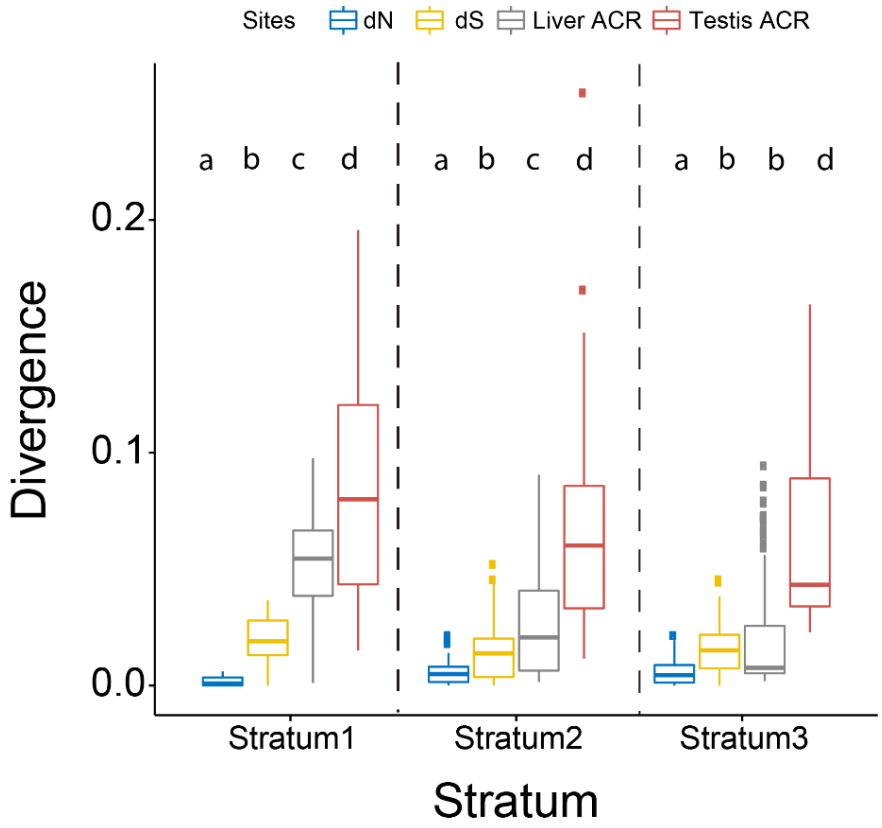
**Declaration of interests**

The authors declare no conflicts of interests.

736 **Figure legends**
737
738 **Figure 1. XY divergence within accessible chromatin regions and coding regions.** Accessible
739 chromatin regions (ACRs) show elevated divergence between the X and Y, relative to nonsynonymous
740 and synonymous substitutions in coding regions. Protein-coding divergence was estimated between the X
741 and Y chromosomes at all shared gametologs. ACRs were identified on the X chromosome and mapped
742 to the Y chromosome. Substitutions within ACRs were summed and divided by the total length of the
743 ACR to calculate divergence. Stratum one: 12 gametologs, 442 liver ACRs, 30 testis ACRs; Stratum two:
744 71 gametologs, 228 liver ACRs, 37 testis ACRs; Stratum three: 46 gametologs, 199 liver ACRs 51 testis
745 ACRs. Rates of sequence divergence for all 4 categories were compared using a Kruskal-Wallis test for
746 each stratum ($P < 0.001$ for all strata). A post-hoc Dunn's test was used to identify which groups were
747 significantly different from one another. P- values were adjusted for multiple comparisons with a
748 Bonferroni correction. The letters above each plot indicate the significant differences between sites within
749 each stratum ($P < 0.001$). Box plots that share a letter assignment are not significantly different from one
750 another ($P > 0.001$).
751
752 **Figure 2 Accessible chromatin divergence between threespine and ninespine stickleback. A.**
753 Predicted substitution rates within accessible regions**.** ACR substitution rates (red) lower than random
754 non-functional control regions (white) and synonymous substitutions (blue) would indicate purifying
755 selection (left). ACR substitution rates roughly equal to non-functional control regions would indicate
756 mutations are accumulating due to inefficient selection from the loss of recombination (middle). ACR
757 substitution rates higher than non-functional control regions would indicate positive selection. **B.** Y-
758 linked ACR divergence is elevated over non-functional intergenic regions. X and Y ACR divergence was
759 compared to randomly drawn non-functional regions matched for GC content (10,000 permutations). The
760 average ACR divergence of each stratum is shown by the dotted line (stratum one: pink, stratum two:
761 green, and stratum three: blue). ACRs were identified on the X chromosome and mapped to the Y
762 chromosome and homologous ninespine stickleback autosome 19 from liver (top) and testis (bottom).
763 Autosome divergence is from chromosome 18. For the sex chromosomes, we identified variants that were
764 unique to the X and Y, by aligning to the outgroup and ignoring fixed differences between species that
765 likely occurred before the sex chromosomes evolved. X- and Y- specific variants within ACRs were
766 summed and divided by the total length of the ACR to calculate divergence. Stratum one: 442 liver ACRs
767 and 30 testis ACRs; stratum two: 228 liver ACRs and 37 testis ACRs; stratum three: 129 liver ACRs and
768 46 testis ACRs; autosomal: 1056 liver ACRs and 305 testis ACRs. Graphic made with BioRender.com
769
770 **Figure 3. Quantifying adaptive divergence and departure from neutrality among accessible**
771 **chromatin regions and coding regions.** Sex-linked ACRs exhibited signatures of positive selection in
772 both liver (left) and testis (middle). Sex-linked coding regions (right) are mostly under purifying selection
773 or not significantly different than neutral expectations. The neutrality index (NI) was calculated using the
774 odds ratio of the McDonald Kreitman test. $NI > 1$ reflects purifying selection (blue) while $NI < 1$ is a
775 signature of positive selection (red). Asterisks denote significant departures from neutrality determined by
776 a chi-square test. Error bars represent two standard deviations calculated from 10,000 bootstrap replicates.
777
778 **Figure 4**. **Gametolog expression level compared to accessible chromatin region divergence.** ACR
779 divergence on the X and Y chromosomes is correlated with X-biased gene expression in both liver and
780 testis tissues. RNA-seq transcript counts that uniquely mapped to the X and Y chromosome were
781 quantified to determine allele-specific expression for all gametologs. Gametolog expression was
782 compared to the average divergence of Y- specific (A and B) and X-specific (C and D) mutations for
783 ACRs within 50 kb of expressed genes. (A). Y liver: $R= 0.40$, $P = 0.004$, $N = 51$; (B.) Y Testis: $R=0.201$,
784 $P = 0.26$, $N = 31$; (C). X Liver: $R = 0.30$, $P = 0.033$, $N = 51$; (D). X Testis: $R = 0.07$, $P = 0.94$, $N = 31$.
785 Shaded regions represent 95% confidence intervals based on bootstrapping. The regressions were not

786 significantly different between the X and Y chromosome for liver (P = 0.590, Fisher transformation) or
787 testis (P = 0.077, Fisher transformation).
788
789
790
791
792 **Figures**

793




794

795 **Figure 1. XY divergence within accessible chromatin regions and coding regions.**

796
797
798
799
800
801
802

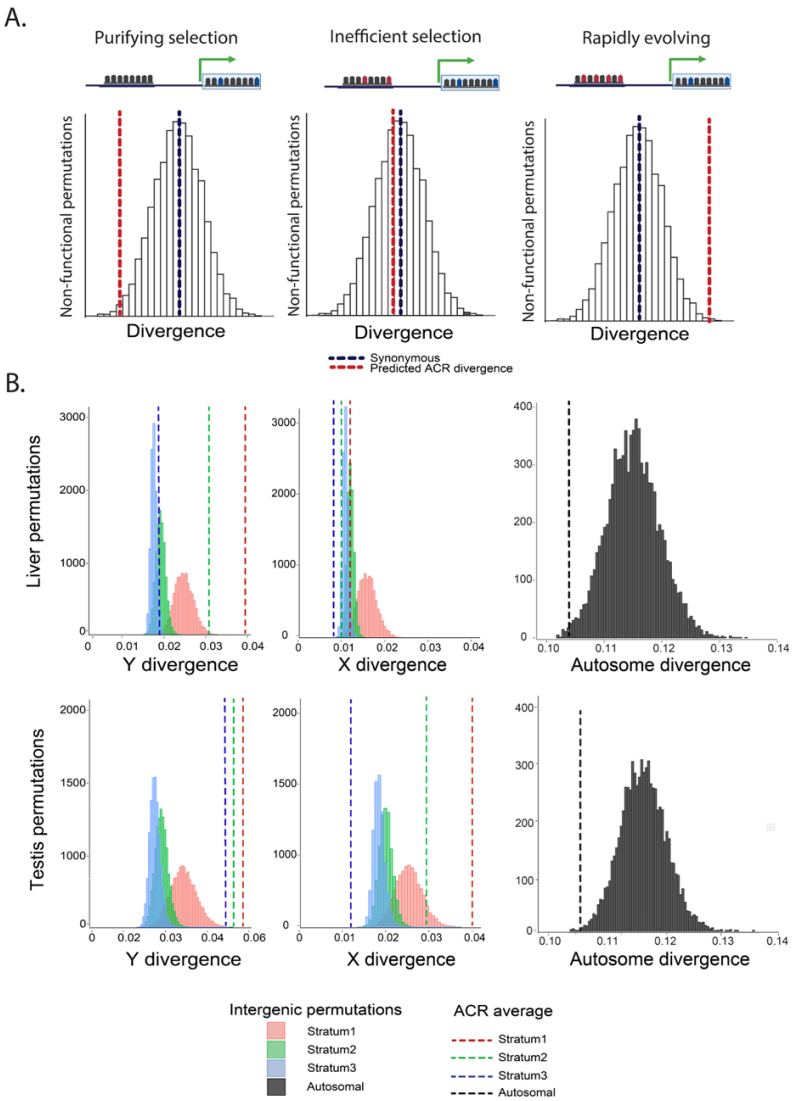
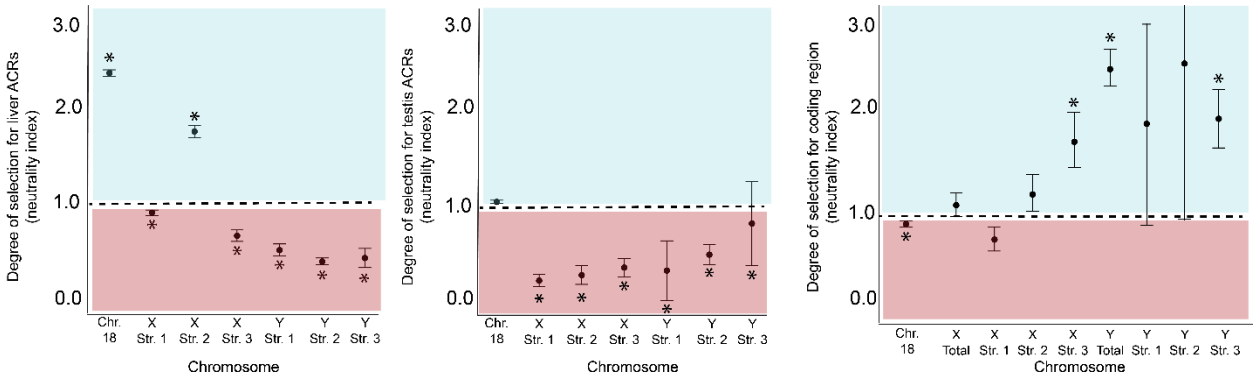



**Figure 2. Accessible chromatin divergence between threespine and ninespine stickleback.**

809



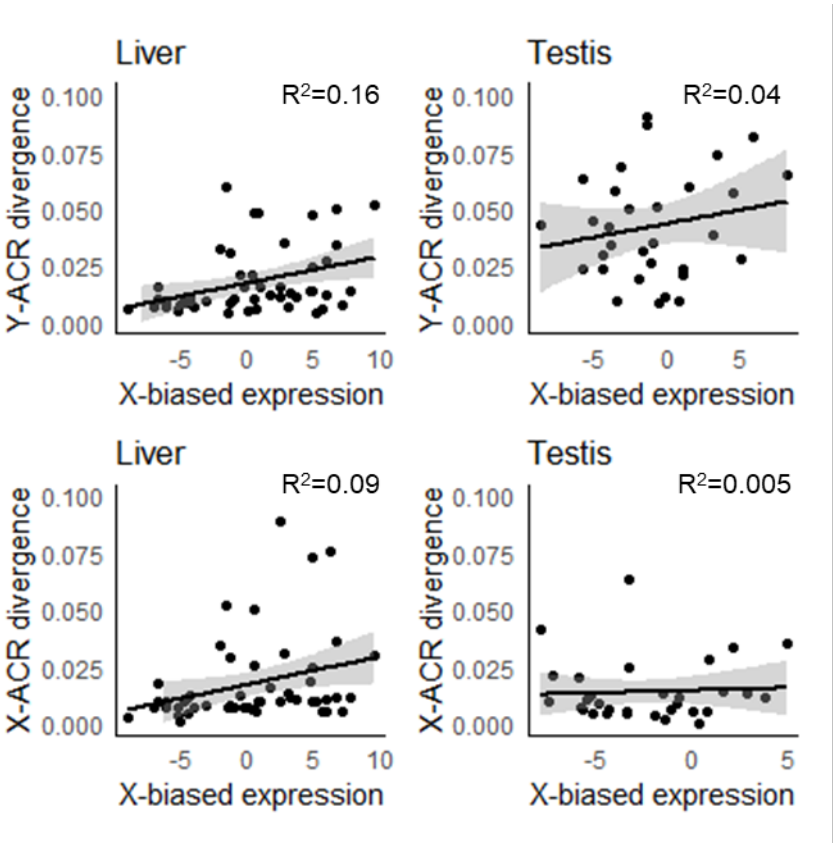

810

811 **Figure 3. Departure from neutrality at sex-linked ACRs.**

812

813

814




815

816 **Figure 4**. **Gametolog expression level compared to ACR divergence.**

817

818

819 **Table 1. McDonald-Kreitman Test for non-coding ACRs.**

| Chromosome | Stratum | Tissue | D(ACR) | D(INT) | P(ACR) | P(INT) | $\alpha$ | NI | ChiSquare | 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 | 89313 | 52716 | 243 | 923 | 0.845 | 0.155 | 870.81 | P<0.001 |
| Y | Stratum 2 | Liver | 30619 | 20743 | 210 | 388 | 0.633 | 0.367 | 147.01 | P<0.001 |
| Y | Stratum 3 | Liver | 16889 | 16875 | 321 | 412 | 0.222 | 0.778 | 11.13 | P<0.001 |
| X | Stratum 1 | Testis | 878 | 760 | 452 | 18109 | 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 | 26658 | 34849 | 6009 | 10003 | 0.215 | 0.785 | 176 | P<0.001 |
| X | Stratum 2 | Liver | 7121 | 13850 | 1265 | 1321 | -0.863 | 1.863 | 224.76 | P<0.001 |
| X | Stratum 3 | Liver | 9644 | 13485 | 2618 | 7618 | 0.519 | 0.481 | 793.25 | P<0.001 |
| Chr18 | NA | Liver | 23380 | 25607 | 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 |

820

821

822

823

824 **Table 2. McDonald-Kreitman Test for coding regions.**

| Chromosome | Subset of coding sites | DN | DS | PN | PS | NI | ChiSquare | P-value |
|---|---|---|---|---|---|---|---|---|
| Y | Whole Chromosome | 16892 | 23964 | 59 | 34 | 2.461 | 18.674 | P<0.001 |
| X | Whole Chromosome | 12586 | 22431 | 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 | 11335 | 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 | 10776 | 135 | 140 | 1.664 | 17.876 | P<0.001 |
| Chr18 | Whole Chromosome | 23795 | 31499 | 2581 | 3823 | 0.890 | 17.485 | P<0.001 |

825

826

827

828

829 **Table 3. Gametologs with Y-biased expression in testes**

| Gene 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 | HMG-box transcription factor 1 |
| ENSGACG00000010672 | roraa | RAR-related orphan receptor A, paralog a |
| ENSGACG00000010898 | NA | NA |
| ENSGACG00000011130 | mrpl23 | mitochondrial ribosomal protein L23 |
| ENSGACG00000011958 | ergic2 | si:ch211-225b10.3 |
| ENSGACG00000012079 | NA | WNK lysine deficient protein kinase 1a |
| ENSGACG00000012796 | NA | NA |
| ENSGACG00000012946 | c2cd5 | C2 calcium dependent domain containing 5 |
| ENSGACG00000013060 | IQSEC3 | IQ motif and Sec7 domain ArfGEF 3b |
| ENSGACG00000013333 | SRGAP1 | SLIT-ROBO Rho GTPase activating protein 1a |
| ENSGACG00000013345 | rxylt1 | ribitol xylosyltransferase 1 |
| ENSGACG00000013388 | mtss1la | MTSS 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 |

## Supplemental Figures

**Supplemental Figure 1. Nucleotide diversity (pi) across the sex chromosomes and representative autosome.** The Y chromosome exhibits a strong reduction in nucleotide diversity consistent with a lower effective population size, and selection. Estimates of nucleotide diversity were calculated from 12 male individuals from a freshwater population. Nucleotide diversity was averaged across 10,000bp windows for A. Chromosome 18, B. X chromosome, and C. Y chromosome. D. The Y chromosome has lower diversity than both the X chromosome and chromosome 18**.** E. Distinct strata on the X chromosome have varying levels of the nucleotide diversity. A Kruskal Wallis test indicated differences between the groups ($P < 0.001$). A post-Hoc Dunn's test and Bonferroni Correction revealed that the PAR was significantly higher from all three strata ($P < 0.001$). Stratum one was significantly higher than stratum two and three ($P < 0.001$). After removing windows that overlapped the centromeric region, stratum two and three are not significantly different ($P = 0.937$). F. All three strata across the Y chromosome have similar estimates of nucleotide diversity ($P=0.768$, Kruskal Wallis test). The PAR was not analyzed for the Y chromosome as the assembly is missing most of the PAR sequence. 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 ($P > 0.001$).

**Supplemental Figure 2. $d_N$ / $d_S$ ratios of coding regions compared to accessible chromatin region divergence.** There is no correlation between $d_N$ / $d_S$ and ACR divergence using a Spearman's Rank Correlation using ACRs from liver (A and B) or testis (C and D) tissue. Liver: (**A**) Y divergence: (Stratum one (Blue): $R=-0.12$ $P = 0.230$, $N = 30$, Stratum two (Yellow): $R = 0.2$, $P = 0.430$, $N=61$;

855 Stratum three (Grey): R = 0.11, P = 0.770 N=126. (**B**) X divergence: (Stratum one (Blue): R = -0.080, P =
856 0.420, N=30; Stratum two (Yellow): R = 0.032, P = 0.9, N = 61; Stratum three (Grey): R = -0.054, P =
857 0.890, N = 126. Similar results in testis: (**C**) Y divergence: (Stratum one (Blue): R = -0.020, P = 0.650,
858 N=14; Stratum two (Yellow): R = -0.430, P = 0.160, N = 12; Stratum three (Grey): R = -0.140, P = 0.480,
859 N = 17. (**D**) X divergence: (Stratum one (Blue): R = 0.520, P = 0.120, N=14; Stratum two (Yellow): R = -
860 0.360, P = 0.210, N = 12; Stratum three (Grey): R = -0.020, P = 0.670, N = 17.
861

862 **Supplemental Figure 3. Selective constraint of genes affects correlation between ACR and proximal**
863 **coding divergence.** Y-linked gametologs were classified by their mode of selection based on a McDonald
864 Kreitman Test. Genes were classified as Positive selection: $NI_{Haldane} < -0.25$, Relaxed purifying selection: -
865 $0.25 < NI_{Haldane} < 0.25$, and purifying selection: $NI_{Haldane} > 0.25$. Genes likely evolving under positive
866 selection had Y-linked and X-linked ACR divergence significantly correlated with $d_N/d_S$ (Y chromosome:
867 R=0.92, P=0.002, N=10; X chromosome: R = 0.81, P = 0.03, Spearman's Rank correlation coefficient).
868 Genes under purifying selection and relaxed purifying selection did not show correlations between ACR
869 divergence and DN/DS (Purifying: Y chromosome: R= -0.14, P = 0.30, N=57; X chromosome: R= -
870 0.022, P =0.87, N=57; Relaxed purifying selection: Y chromosome: R= -0.293, P =0.0155, N=35, X
871 chromosome: R= -0.32, P =0.063, N=35, Spearman's Rank Correlation coefficient. Grey areas represent
872 95% confidence intervals. Outliers with $d_N/d_S > 3$ were removed prior to analysis. When $d_N/d_S > 1$ are
873 removed the relationship for genes under positive selection is still significant (R=0.86, P=0.006).

874 **Supplemental figure 4. Rates of allelic expression for Y gametologs under different modes of**
875 **selection.** Y-linked gametologs were classified by their mode of selection based on a McDonald Kreitman
876 Test. Genes were classified as Positive selection: $NI_{Haldane} < -0.25$, Relaxed constraint selection: -0.25
877 $< NI_{Haldane} < 0.25$, and purifying selection: $NI_{Haldane} > 0.25$. X/Y gene expression was quantified based on
878 counts from liver and testis RNA-seq data from 3 individuals. Counts were averaged across all three
879 replicates. Rates of X/Y expression are not significantly different for each class (Kruskal-Wallis test).

880

881 **Supplemental Figure 5. Accessible chromatin region divergence compared to functional states of Y-**
882 **linked gametologs.** ACR divergence was not significantly different among gametologs that contained
883 mutations that potentially impede function (i.e., premature stop codons, frameshifts, deleted nucleotides)
884 or are functional. Y-specific functional mutations were identified by comparing coding sequence between
885 the Y chromosome and the outgroup autosomal coding sequence in the ninespine stickleback. ACRs were
886 assigned to the closest gametolog within 50kb. P > 0.05 for all pairwise comparisons for both liver (A and
887 C) and testis (B and D) for regulatory divergence on the X and Y.
888
889

890 **Supplemental Figure 6. High concordance of chromatin accessibility around transcription start site**
891 **of expressed genes**. Transcription start sites (TSSs) for all genes were ordered by RNA-seq read depth.
892 ATAC-seq read depth was quantified ±3 kb of TSSs. Similar ATAC-seq read depth was observed
893 between replicates. ATAC-seq depth across replicates is a strong predictor of expression from RNA-seq
894 from same tissue type.
895

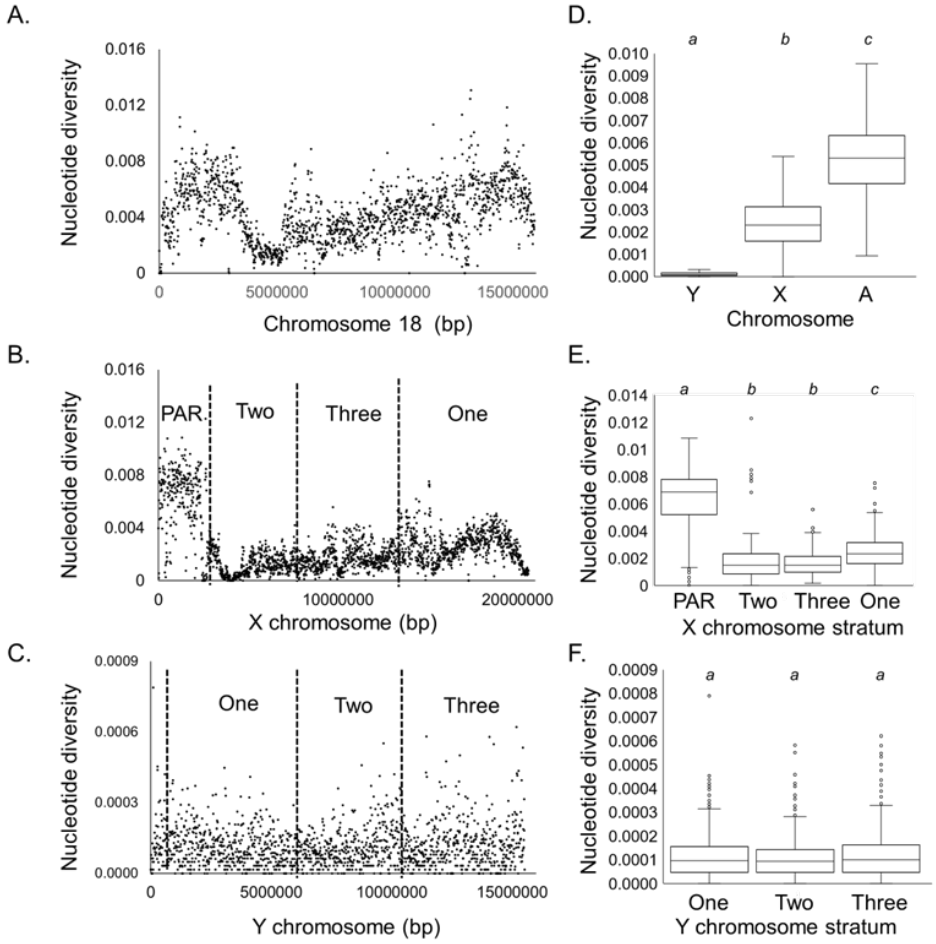
896 **Supplemental Figure 7. High concordance of chromatin accessibility between peaks identified in**
897 **two replicates.** Peaks were identified in two replicates of liver (left) and testis (right). $Log_{-10}$(P-Value) of
898 each shared peak is highly correlated between replicates. Irreproducible Discovery Rate was calculated by
899 calculated the ranked significance of reproducible and irreproducible peaks between replicates. Peaks
900 with an IDR < 0.05 are shown in red.
901

902 **Supplemental Figure 8. Verification of ATAC-seq proximity to annotated genes.** Three
903 representative examples of ATAC-seq read depth enriched at ACRs called with MACS2. The most
904 proximal genes were identified with homer and with a custom python script. Annotations for a subset of
905 100 X chromosome ACR were visually inspected with JBrowse, to confirm that the peak was accurately
906 assigned to nearby genes. ATAC-seq read depth is shown in grey. Bed files for the summit of each ACR
907 identified by MACS2 is shown below the Ensembl gene annotations.

908

909

910

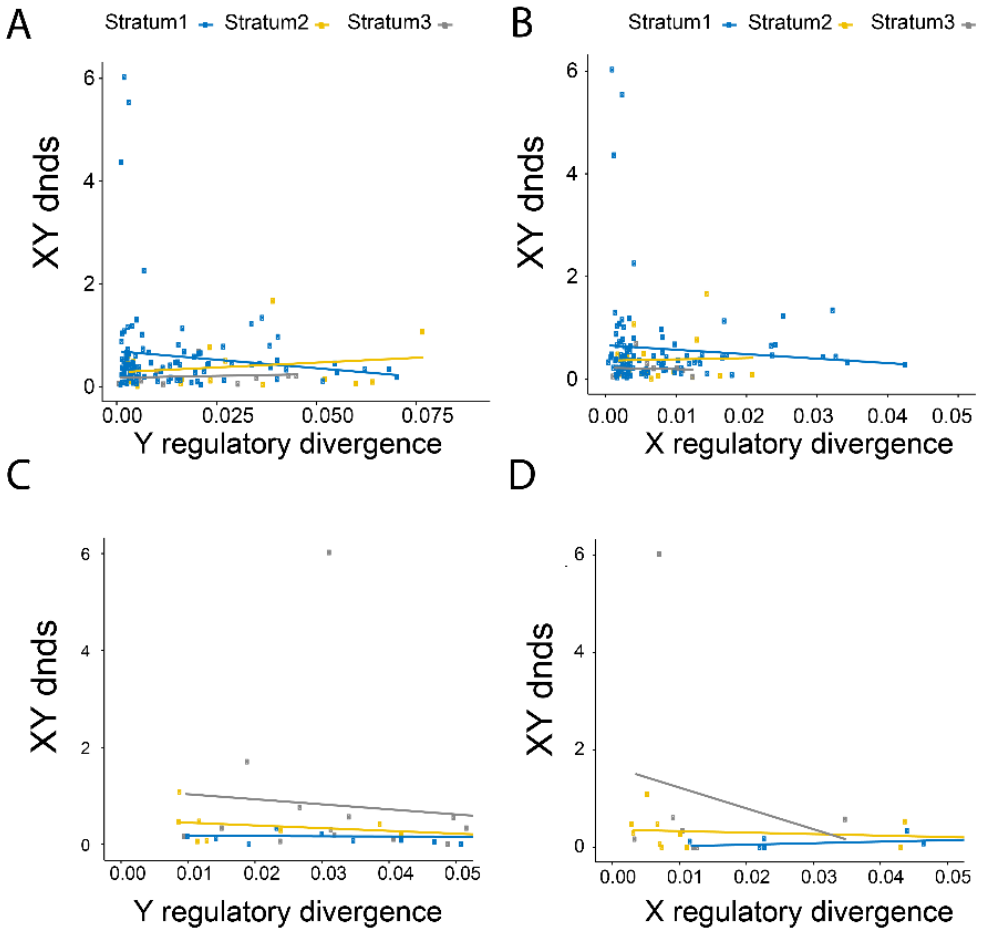



911

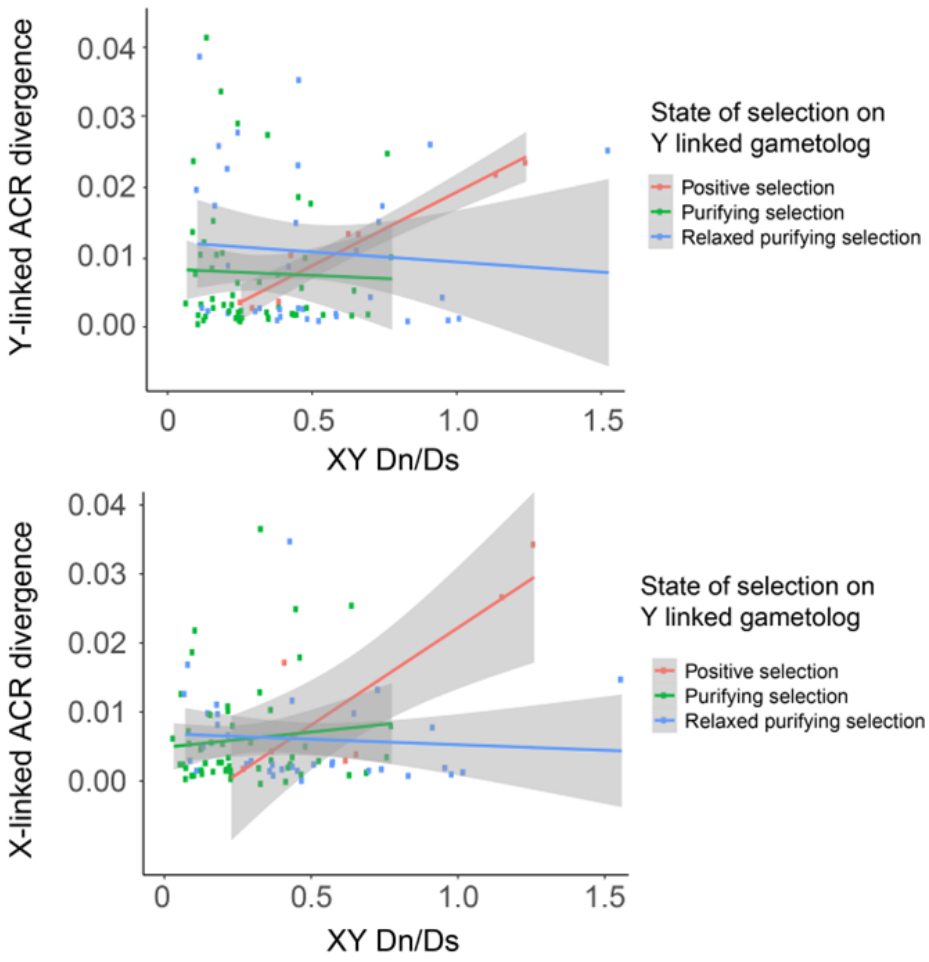
912 **Supplemental Figure 1. Nucleotide diversity (pi) across the sex chromosomes and representative**
913 **autosome.**

914

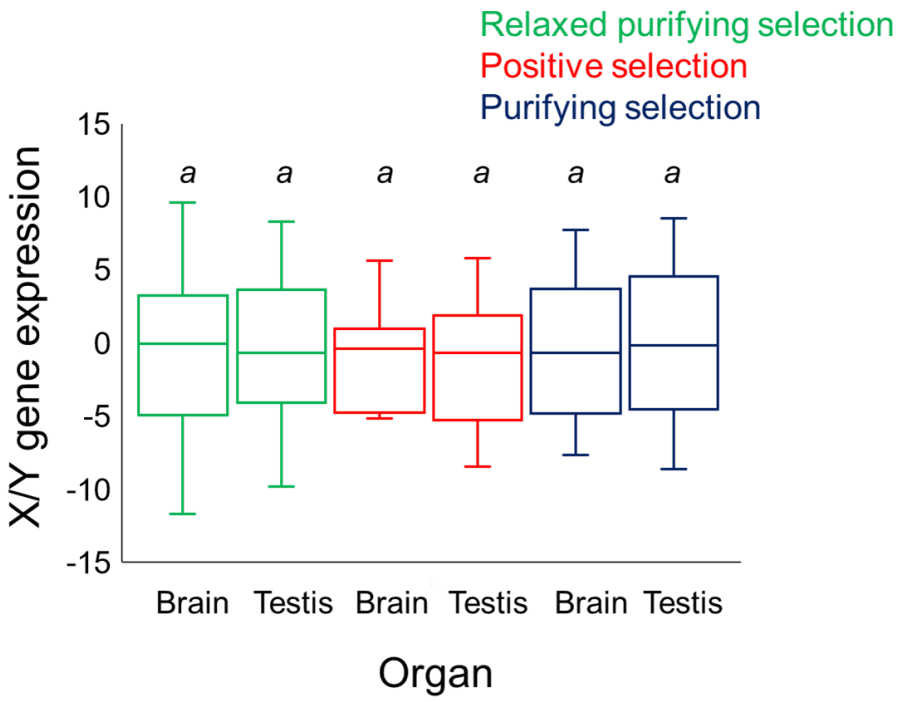
915




**Supplemental Figure 2. $d_N$ / $d_S$ ratios of coding regions compared to ACR divergence.**

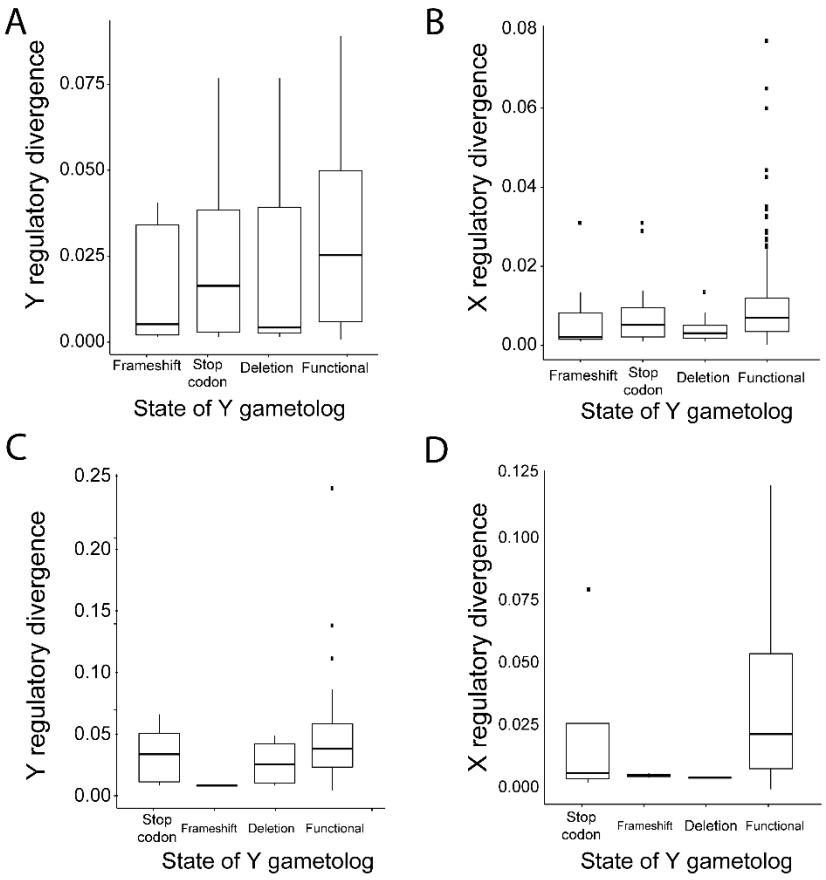



**Supplemental Figure 3. Selective constraint of genes affects correlation between ACR and proximal coding divergence.**




930

931 **Supplemental figure 4. Allelic-specific expression for Y gametologs under different modes of**
932 **selection**

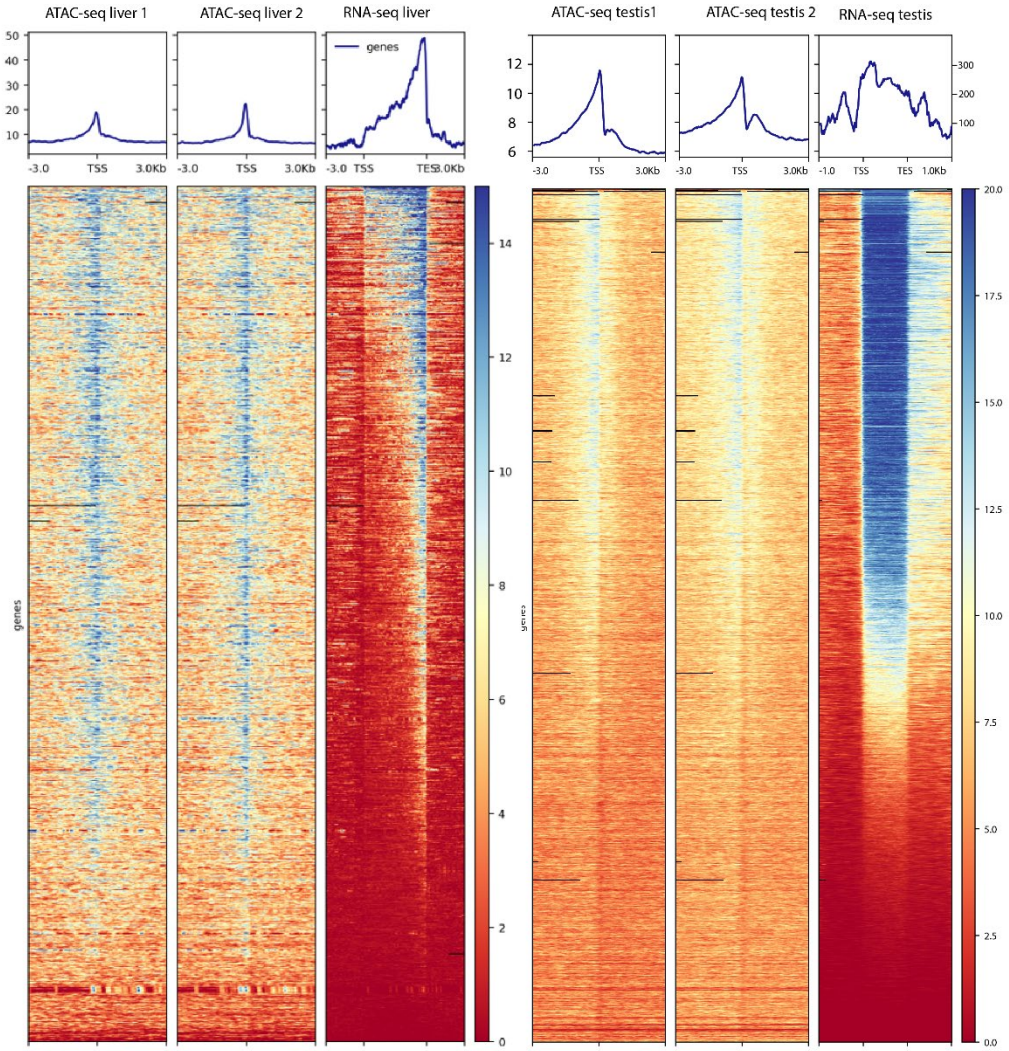



933

934 **Supplemental Figure 5. ACR divergence compared to functional states of Y-linked gametologs.**

935

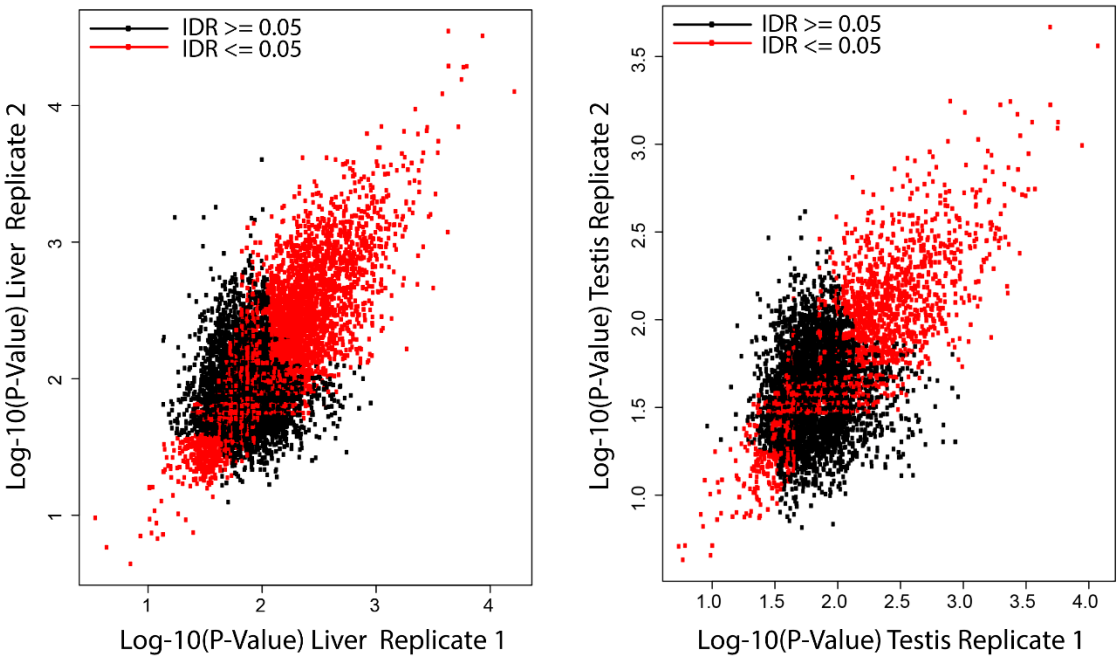
936




937

938 **Supplemental Figure 6. High concordance of chromatin accessibility around transcription start site**
939 **of expressed genes.**

940

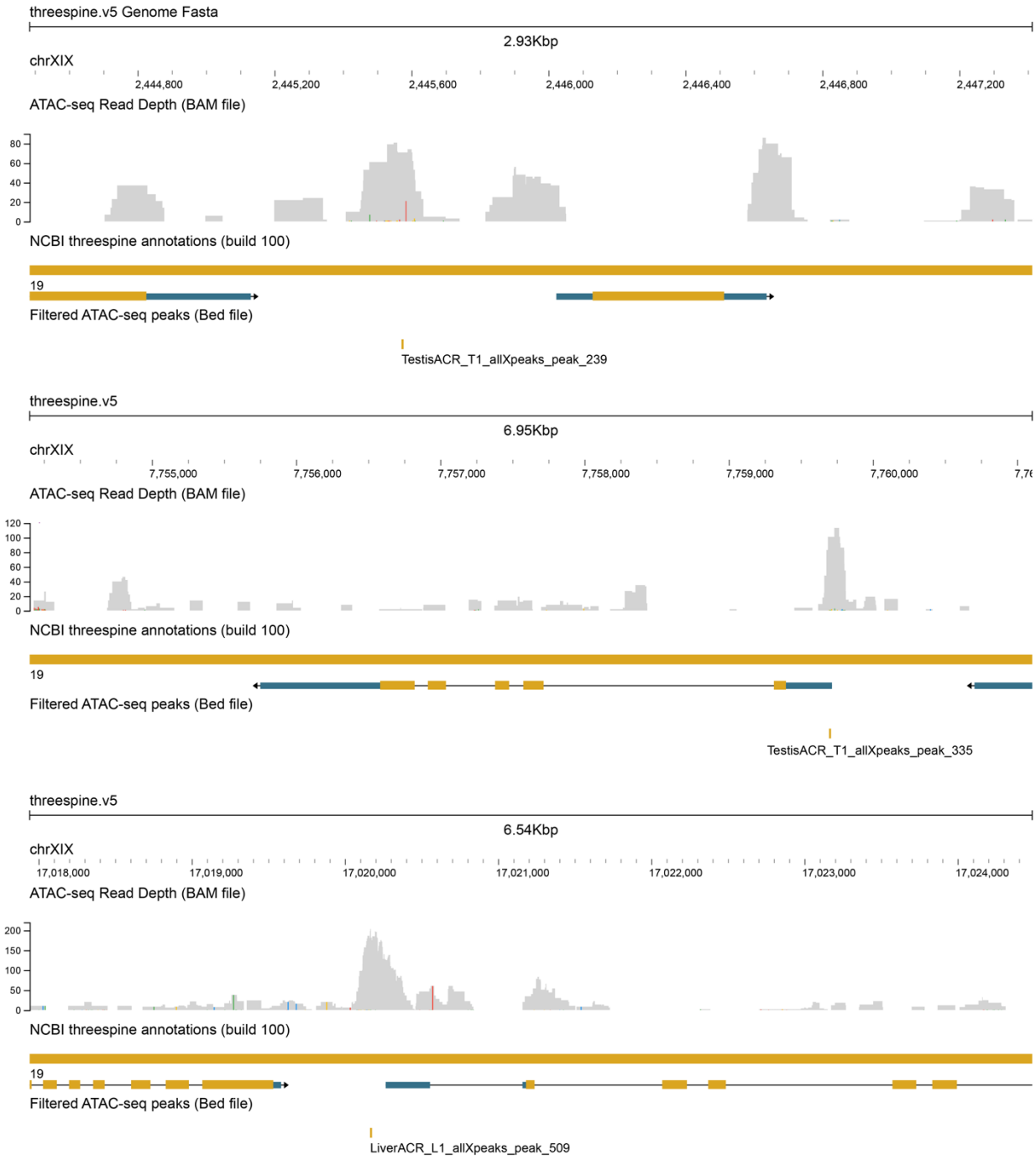



941

942 **Supplemental Figure 7. High concordance of chromatin accessibility between peaks identified in**

943 **two replicates.**

944




**Supplemental Figure 8. Verification of ATAC-seq proximity to annotated genes.**

950 **References**
951


952 Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, Sullivan AM, Bemm F,
953 Jolic D, Arsovski AA, et al. 2018. Complex Relationships between Chromatin Accessibility,
954 Sequence Divergence, and Gene Expression in Arabidopsis thaliana. Molecular Biology and
955 Evolution 35:837-854.
956 Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing
957 data. Bioinformatics 31:166-169.
958 Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149-1152.
959 Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. 2014. Quantitative
960 genome-wide enhancer activity maps for five Drosophila species show functional enhancer
961 conservation and turnover during cis-regulatory evolution. Nature Genetics 46:685-692.
962 Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. 2013. RNA
963 sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken
964 and allows comprehensive annotation of the W-chromosome. Genome Biology 14:R26.
965 Bachtrog D. 2006. Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. Current
966 Biology 16:1694-1699.
967 Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in
968 Drosophila. Genome Research 15:1393-1401.
969 Bachtrog D. 2008. The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration.
970 Genetics 179:1513-1525.
971 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome
972 degeneration. Nature Reviews Genetics 14:113-124.
973 Bachtrog D, Hom E, Wong KM, Maside X, De Jong P. 2008. Genomic degradation of a young Y
974 chromosome in Drosophila miranda. Genome Biology 9:R30.
975 Bachtrog D, Jensen JD, Zhang Z. 2009. Accelerated Adaptive Evolution on a Newly Formed X
976 Chromosome. PLoS Biology 7:e1000082.
977 Bachtrog D, Mahajan S, Bracewell R. 2019. Massive gene amplification on a recently formed Drosophila
978 Y chromosome. Nature Ecology & Evolution 3:1587-1597.
979 Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I,
980 Ming R, et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLoS Biology
981 12:e1001899.
982 Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008. Effects of X-Linkage and Sex-Biased Gene Expression on
983 the Rate of Adaptive Protein Evolution in Drosophila. Molecular Biology and Evolution 25:1639-
984 1650.
985 Beaudry FEG, Barrett SCH, Wright SI. 2017. Genomic Loss and Silencing on the Y Chromosomes of
986 Rumex. Genome Biology and Evolution 9:3345-3355.
987 Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T,
988 Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive
989 regulators. Nature 508:494-499.
990 Bellott DW, Page DC. 2021. Dosage-sensitive functions in embryonic development drove the survival of
991 genes on sex-specific chromosomes in snakes, birds, and mammals. Genome Research 31:198-
992 210.
993 Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, Galkina S, Pyntikova T, Koutseva N, Graves T,
994 et al. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive
995 regulators. Nature Genetics 49:387-394.

996 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
997 Bioinformatics 30:2114-2120.
998 Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A,
999 Kircher M. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343-
1000 348.
1001 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for
1002 fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and
1003 nucleosome position. Nature Methods 10:1213-1218.
1004 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
1005 architecture and applications. BMC Bioinformatics 10:421.
1006 Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA. 2008. Evolution of Primate Gene Expression: Drift
1007 and Corrective Sweeps? Genetics 180:1379-1389.
1008 Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. 2022. Unique structure and positive
1009 selection promote the rapid divergence of Drosophila Y chromosomes. eLife 11.
1010 Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. Proceedings of
1011 the National Academy of Sciences 75:5618-5622.
1012 Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philosophical Transactions
1013 of the Royal Society of London. Series B: Biological Sciences 355:1563-1572.
1014 Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and
1015 autosomes. The American Naturalist 130:113-146.
1016 Connelly CF, Wakefield J, Akey JM. 2014. Evolution and Genetic Architecture of Chromatin Accessibility
1017 and Function in Yeast. PLoS Genetics 10:e1004427.
1018 Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. 2015. Molecular mechanisms
1019 and evolutionary processes contributing to accelerated divergence of gene expression on the
1020 Drosophila X chromosome. Molecular Biology and Evolution 32:2605-2615.
1021 Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. 2013. Analysis of expressed SNPs identifies
1022 variable extents of expression from the human inactive X chromosome. Genome Biology
1023 14:R122.
1024 Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, Austine-Orimoloye O,
1025 Azov Andrey G, Barnes I, Bennett R, et al. 2021. Ensembl 2022. Nucleic Acids Research 50:D988-
1026 D995.
1027 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT,
1028 Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.
1029 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
1030 Acids Research 32:1792-1797.
1031 Ellison C, Bachtrog D. 2019. Recurrent gene co-amplification on Drosophila X and Y chromosomes. PLoS
1032 Genetics 15:e1008251.
1033 Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R,
1034 Fischer A, et al. 2013. Comparative Population Genomics of the Ejaculate in Humans and the
1035 Great Apes. Molecular Biology and Evolution 30:964-976.
1036 Haldane J. 1956. The estimation and significance of the logarithm of a ratio of frequencies. Annals of
1037 human genetics 20:309-311.
1038 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010.
1039 Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory
1040 Elements Required for Macrophage and B Cell Identities. Molecular Cell 38:576-589.
1041 Horvath R, Josephs EB, Pesquet E, Stinchcombe JR, Wright SI, Scofield D, Slotte T. 2021. Selection on
1042 Accessible Chromatin Regions in Capsella grandiflora. Molecular Biology and Evolution 38:5563-
1043 5575.

1044 Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki
1045 C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y
1046 chromosomes. Nature 483:82-86.
1047 Hughes JF, Skaletsky H, Pyntikova T, Graves TA, Van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke
1048 DP, Friedman C, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent
1049 in structure and gene content. Nature 463:536-539.
1050 Hughes JF, Skaletsky H, Pyntikova T, Koutseva N, Raudsepp T, Brown LG, Bellott DW, Cho T-J, Dugan-
1051 Rocha S, Khan Z, et al. 2020. Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms
1052 races in mammalian lineages. Genome Research 30:1716-1726.
1053 Jalili V, Matteucci M, Masseroli M, Morelli MJ. 2015. Using combined evidence from replicates to
1054 evaluate ChIP-seq peaks. Bioinformatics 31:2761-2769.
1055 Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S.
1056 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and
1057 chimpanzees. Science 309:1850-1854.
1058 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of
1059 transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology
1060 14:R36.
1061 Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Schartl M. 2021. Evolution of the canonical sex
1062 chromosomes of the guppy and its relatives. G3 Genes|Genomes|Genetics.
1063 Kopania EEK, Larson EL, Callahan C, Keeble S, Good JM. 2022. Molecular Evolution across Mouse
1064 Spermatogenesis. Molecular Biology and Evolution 39.
1065 Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory cis-trans
1066 Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila.
1067 Genetics 171:1813-1822.
1068 Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB,
1069 Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE
1070 consortia. Genome Research 22:1813-1831.
1071 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
1072 Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BA, Smith AD, Dean MD, Good JM. 2016. Contrasting
1073 Levels of Molecular Evolution on the Mouse X Chromosome. Genetics 203:1841-1857.
1074 Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong Purifying Selection at Synonymous Sites in
1075 D. melanogaster. PLoS Genetics 9:e1003527.
1076 Lawson Handley LJ, Berset-BräNdli L, Perrin N. 2006. Disentangling Reasons for Low Y Chromosome
1077 Variation in the Greater White-Toothed Shrew (Crocidura russula). Genetics 173:935-942.
1078 Lenormand T, Fyon F, Sun E, Roze D. 2020. Sex Chromosome Degeneration by Regulatory Evolution.
1079 Current Biology 30:3001-3006.e3005.
1080 Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual
1081 dimorphism. Science 375:663-666.
1082 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The
1083 Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
1084 Link V, Aguilar-Gómez D, Ramírez-Suástegui C, Hurst LD, Cortez D. 2017. Male Mutation Bias Is the Main
1085 Force Shaping Chromosomal Substitution Rates in Monotreme Mammals. Genome Biology and
1086 Evolution 9:2198-2210.
1087 Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. 2017. Combining ATAC-seq with nuclei sorting
1088 for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Res 45:e41.
1089 Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. 2019. The prevalence, evolution and
1090 chromatin signatures of plant regulatory elements. Nature Plants 5:1250-1259.

1091 Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. Nature
1092 Communications 8.
1093 Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young Drosophila Y
1094 chromosome using single-molecule sequencing and chromatin conformation capture. PLoS
1095 Biology 16:e2006348.
1096 Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GAB, Touzet P. 2019. Evolution of
1097 Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination
1098 Systems. Genome Biology and Evolution 11:350-361.
1099 Martínez-Pacheco M, Tenorio M, Almonte L, Fajardo V, Godínez A, Fernández D, Cornejo-Páramo P,
1100 Díaz-Barba K, Halbert J, Liechti A, et al. 2020. Expression Evolution of Ancestral XY Gametologs
1101 across All Major Groups of Placental Mammals. Genome Biology and Evolution 12:2015-2028.
1102 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends in Genetics
1103 29:537-544.
1104 Meisel RP, Malone JH, Clark AG. 2012a. Disentangling the relationship between sex-biased gene
1105 expression and X-linkage. Genome Research 22:1255-1265.
1106 Meisel RP, Malone JH, Clark AG. 2012b. Faster-X Evolution of Gene Expression in Drosophila. PLoS
1107 Genetics 8:e1003013.
1108 Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey
1109 W, Ewels PA, et al. 2015. Mapping long-range promoter contacts in human cells with high-
1110 resolution capture Hi-C. Nature Genetics 47:598-606.
1111 Muller HJ. 1918. Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal
1112 Factors. Genetics 3:422-499.
1113 Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid De Novo Evolution of X
1114 Chromosome Dosage Compensation in Silene latifolia, a Plant with Young Sex Chromosomes.
1115 PLoS Biology 10:e1001308.
1116 Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Hobza R, Picard F, Widmer A,
1117 et al. 2018. Genomic imprinting mediates dosage compensation in a young plant XY system.
1118 Nature Plants 4:677-680.
1119 Naftaly AS, Pau S, White MA. 2021. Long-read RNA sequencing reveals widespread sex-specific
1120 alternative splicing in threespine stickleback fish. Genome Research:gr.274282.274120.
1121 Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using
1122 long-read sequencing. G3 Genes|Genomes|Genetics 11.
1123 Nesta AV, Tafur D, Beck CR. 2021. Hotspots of Human Mutation. Trends in Genetics 37:717-729.
1124 Orr HA, Kim Y. 1998. An Adaptive Hypothesis for the Evolution of the Y Chromosome. Genetics
1125 150:1693-1698.
1126 Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient
1127 extraction of SNPs from multi-FASTA alignments. Microbial Genomics 2.
1128 Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression.
1129 Nature Reviews Genetics 14:83-87.
1130 Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM,
1131 Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals
1132 convergent signatures of sex chromosome evolution. Genome Biology 21.
1133 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
1134 Bioinformatics 26:841-842.
1135 Qunhua L, James BB, Haiyan H, Peter JB. 2011. Measuring reproducibility of high-throughput
1136 experiments. The Annals of Applied Statistics 5:1752-1779.

1137 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016.
1138 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids
1139 Research 44:W160-W165.
1140 Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among
1141 genes from Drosophila, mice, and humans. Molecular Biology and Evolution 13:735-748.
1142 Reske JJ, Wilson MR, Chandler RL. 2020. ATAC-seq normalization method can significantly affect
1143 differential accessibility analysis and interpretation. Epigenetics & Chromatin 13.
1144 Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-
1145 Tatché M, et al. 2019. Widespread long-range cis-regulatory elements in the maize genome.
1146 Nature Plants 5:1237-1249.
1147 Rodríguez Lorenzo JL, Hobza R, Vyskot B. 2018. DNA methylation and genetic degeneration of the Y
1148 chromosome in the dioecious plant Silene latifolia. BMC Genomics 19.
1149 Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome—patterns and
1150 consequences. Molecular Ecology 22:3014-3027.
1151 Rollins JL, Chiang P, Waite JN, Von Hippel FA, Bell MA. 2017. Jacks and jills: Alternative life-history
1152 phenotypes and skewed sex ratio in anadromous threespine stickleback (Gasterosteus
1153 aculeatus). Evolutionary Ecology Research 18:363-382.
1154 Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. 2021. Heterogeneous Histories of
1155 Recombination Suppression on Stickleback Sex Chromosomes. Molecular Biology and Evolution
1156 38:4403-4418.
1157 Schoenfelder S, Fraser P. 2019. Long-range enhancer–promoter contacts in gene expression control.
1158 Nature Reviews Genetics 20:437-455.
1159 Shanfelter AF, Archambeault SL, White MA. 2019. Divergent Fine-Scale Recombination Landscapes
1160 between a Freshwater and Marine Population of Threespine Stickleback Fish. Genome Biology
1161 and Evolution 11:1552-1572.
1162 Shaw DE, White MA. 2022. The evolution of gene regulation on sex chromosomes. Trends in Genetics.
1163 Signor SA, Nuzhdin SV. 2018. The Evolution of Gene Expression in cis and trans. Trends in Genetics
1164 34:532-544.
1165 Singh ND, Koerich LB, Carvalho AB, Clark AG. 2014. Positive and Purifying Selection on the Drosophila Y
1166 Chromosome. Molecular Biology and Evolution 31:2612-2623.
1167 Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J,
1168 Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete
1169 sequence classes. Nature 423:825-837.
1170 Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC
1171 Bioinformatics 6:31.
1172 Soh YQ, Alföldi J, Pyntikova T, Laura, Graves T, Patrick, Robert, Kremitzki C, Koutseva N, Jacob, et al.
1173 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and
1174 Amplification on Both Sex Chromosomes. Cell 159:800-813.
1175 Stoletzki N, Eyre-Walker A. 2010. Estimation of the Neutrality Index. Molecular Biology and Evolution
1176 28:63-70.
1177 Tomaszkiewicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D, O'Brien PCM,
1178 Chikhi R, Ryder OA, et al. 2016. A time- and cost-effective strategy to sequence mammalian Y
1179 Chromosomes: an application to the de novo assembly of gorilla Y. Genome Research 26:530-
1180 540.
1181 Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A,
1182 et al. 2017. Landscape of X chromosome inactivation across human tissues. Nature 550:244-248.

1183 Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS,
1184 Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius pungitius)
1185 genome. Genome Biology and Evolution.
1186 Veltsos P, Ridout KE, Toups MA, González-Martínez SC, Muyle A, Emery O, Rastas P, Hudzieczek V, Hobza
1187 R, Vyskot B, et al. 2019. Early Sex-Chromosome Evolution in the Diploid Dioecious Plant
1188 Mercurialis annua. Genetics 212:815-835.
1189 Vermunt MW, Tan SC, Castelijns B, Geeven G, Reinink P, De Bruijn E, Kondova I, Persengiev S, Bontrop R,
1190 Cuppen E, et al. 2016. Epigenomic annotation of gene regulatory alterations during evolution of
1191 the primate brain. Nature Neuroscience 19:494-503.
1192 Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model.
1193 Evolution: International Journal of Organic Evolution 63:2413-2426.
1194 Wei KHC, Bachtrog D. 2019. Ancestral male recombination in Drosophila albomicans produced
1195 geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. PLoS
1196 Genetics 15:e1008502.
1197 White MA, Kitano J, Peichel CL. 2015. Purifying Selection Maintains Dosage-Sensitive Genes during
1198 Degeneration of the Threespine Stickleback Y Chromosome. Molecular Biology and Evolution
1199 32:1981-1995.
1200 Wilson Sayres MA. 2018. Genetic Diversity on the Sex Chromosomes. Genome Biology and Evolution
1201 10:1064-1078.
1202 Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural Selection Reduced Diversity on Human Y
1203 Chromosomes. PLoS Genetics 10:e1004064.
1204 Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data
1205 analysis. Genome Biology 21.
1206 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li
1207 W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9:R137.

1208

1209

1210