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 ideal 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 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 show 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 as a result of sequence degeneration through the accumulation of deleterious mutations, growing evidence indicates not all genes can be 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 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 spaces. 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 techniques, like ATAC-seq (Buenrostro, et al. 2013), utilize short-read sequencing to profile accessible chromatin regions (ACRs) at a fine-scale across the genome. Across autosomes, ACRs show signatures of purifying selection relative to other intergenic regions, consistent with these regions containing important

79 functional elements for gene regulation (Connelly, et al. 2014; Lu, et al. 2019; Horvath, et al. 2021).
80 Profiling ACRs on sex chromosomes would therefore provide a means to study the molecular evolution
81 of *cis*-regulatory regions in the context of Y degeneration.

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

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

109

110 **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. Accessible chromatin regions (ACRs) are sites of the genome that 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 Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) from liver tissue (Naftaly, et al. 2021) and testis tissue. We identified X-linked ACRs and used these as the set of ancestral ACRs that were likely present in the shared autosome ancestor before the Y chromosome began degenerating. 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 orthologous regions on the Y chromosome and the ancestral ninespine stickleback *(Pungititus pungitius)* autosome 19. In testis tissue, 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.

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 non-synonymous mutations among all three strata (Figure 1) ($P < 0.001$; Kruskal-Wallis and Dunn's Test). We also compared ACR divergence with synonymous substitutions, which is used as measure of neutral evolution. Testis ACRs had divergence higher than neutral, synonymous substitutions in all three strata (Figure 1) ($P < 0.001$; Kruskal-Wallis and Dunn's Test). Liver ACR's exhibited divergence that was higher than synonymous substitutions in the two oldest strata (one and two) ($P < 0.001$, Kruskal-

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

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

168 ACRs functional in the liver exhibited higher substitution rates on the Y chromosome than on the
169 X chromosome. Y-linked ACRs had significantly higher substitution rates compared to the non-functional
170 control regions for the two oldest evolutionary strata (Figure 2B, P < 0.001; stratum one and two, 10,000
171 permutations). In contrast to the Y-linked ACRs, we found the X-linked liver ACR sequence divergence
172 was significantly lower than non-functional control regions, consistent with purifying selection operating
173 across all three strata (Figure 2B; Stratum one: P = 0.016; Stratum two and three: P < 0.001; 10,000
174 permutations). This suggests that ACRs may be functionally constrained more on the X chromosome,
175 compared to Y-linked ACRs. We observed a similar pattern on the Y chromosome for ACRs function in
176 testis tissue. Y-linked ACRs had significantly higher substitution rates in all three evolutionary strata,

177 compared to non-functional control regions (Figure 2B; P < 0.001, 10,000 permutations). However, we
178 observed X-linked ACRs from testis tissue, also exhibited elevated nucleotide substitutions in the oldest
179 two evolutionary strata (Figure 2B; P < 0.001, 10,000 permutations), while the youngest stratum was
180 significantly lower than intergenic regions (Figure 2B; P < 0.001, 10,000 permutations). This indicates
181 some *cis*-regulatory elements may also be under positive selection on the X chromosome and this occurs
182 in a tissue-specific fashion.

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

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

190 We used short-read sequencing of 12 males from a freshwater population of stickleback fish to
191 search for signatures of selection within ACRs. In the absence of selection, the ratio of nucleotide
192 diversity to sequence divergence should be equal between functional and non-functional sites. Deviation
193 from this expectation would imply these regions are under selection. To explore this, we performed a
194 modified McDonald-Kreitman test on non-coding sequences (Andolfatto 2005). We compared the
195 number of X- and Y- specific substitutions to the number of polymorphisms segregating within a
196 population for ACRs and non-functional control sequences. Except for stratum two on the X
197 chromosome, sex chromosome ACRs from the liver and the testis had a greater proportion of sites fixed
198 by positive selection (α), compared to autosomal ACRs (P < 0.001) (Table 1). We also separately
199 identified the direction of selection acting on ACRs by calculating the odds-ratio of the McDonald
200 Kreitman test, known as the neutrality index (NI) (Rand and Kann 1996). All liver ACRs of the sex
201 chromosomes had an NI < 1, consistent with positive selection (Figure 3, Table 1). The only exception we
202 found was for one of the younger strata on the X chromosome (stratum two), which had elevated NI,
203 suggesting that purifying selection is operating for some X-linked ACRs active in the liver. Compared to
204 the X chromosome and the autosome, the Y chromosome had significantly lower NI (Stratum one and
205 two; P < 0.001). For the testis ACRs, NI was indistinguishable between the X and Y chromosome, and
206 both sex chromosomes were lower than autosomal ACRs (Figure 3) (P < 0.001), indicating positive
207 selection.

208 Among the non-functional control intergenic sites, we detected major differences in nucleotide
209 diversity among the autosomes and sex chromosomes. Due to a lower effective population size, the Y
210 chromosome is expected to have one-quarter of the nucleotide diversity observed on autosomes, given an
211 equal sex ratio in the population. We found Y chromosome nucleotide diversity within intergenic regions
212 was much lower than expected, compared to autosomes (Y chromosome pi: 0.00013; autosome pi:
213 0.0050; Supplemental Figure 1). This suggests that selection acting on linked deleterious or beneficial
214 mutations has lowered nucleotide diversity beyond neutral expectations on the Y chromosome.

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

216 Selection should favor the loss of expression from the Y chromosome in order to silence coding
217 regions that have accumulated deleterious mutations (Orr and Kim 1998; Bachtrog 2006; Lenormand, et
218 al. 2020). To test if Y-linked *cis*-regulatory divergence was associated with changes in expression, we
219 compared ACR sequence divergence to changes in allelic expression on the X and Y from liver and testis
220 RNA-seq transcriptomes (Peichel, et al. 2020). To maximize the total number of genes for this analysis,
221 we pooled genes across all three evolutionary strata. We found that sequence divergence among liver
222 ACRs that were proximal to genes was a predictor of allele-specific expression. We observed higher
223 expression of the X-linked gametolog, relative to the Y-linked gametolog, when ACRs on the Y
224 chromosome had more substitutions (Figure 4A; N = 51*;* liver expression, P = 0.004, Spearman's rank
225 correlation). There was a weaker correlation between expression and substitutions within ACRs on the X
226 chromosome (Figure 4C: N = 51, P = 0.033, Spearman's rank correlation). Although we do not have gene
227 expression from the ninespine stickleback to determine which sex chromosome is deviating from the
228 ancestral expression level, the Y-linked substitutions indicate the X-biased expression we observe is
229 likely due to downregulating the Y chromosome. A similar trend was observed in testis tissue, but the
230 correlations were not significant, likely due to the small number of genes available for the analysis
231 (Figure 4B N = 31, P = 0.111, Spearman's rank correlation).

232 ***Cis*-regulatory evolution is not correlated with coding sequence evolution**

233 The downregulation of Y-linked gametologs could be adaptive if loss of expression follows the
234 accumulation of deleterious coding substitutions (Orr and Kim 1998). We searched for evidence of
235 adaptive silencing on the Y chromosome by comparing the ACR divergence to the ratio of non-
236 synonymous to synonymous substitutions (d_N/d_S) within the coding region of each gene. An adaptive
237 silencing model would be supported if Y-linked gametologs with an elevated d_N/d_S ratio also have
238 elevated *cis*-regulatory divergence. Among genes shared between the X and Y chromosomes, we found
239 no correlation between ACR nucleotide divergence and d_N/d_S (Supplemental Figure 2; each stratum on X

240 and Y: P > 0.05; Spearman's rank correlation), indicating coding regions with elevated ACR divergence
241 are not more likely to have elevated substitutions in the amino acid sequence. We also examined whether
242 ACR divergence was higher within genes that contained frameshift mutations, nonsense mutations, or in-
243 frame deletions. These types of mutations are more likely to produce a non-functional or sub-optimal
244 peptide. We found no significant difference in the number of ACR substitutions from liver and testis
245 tissue for coding regions that contained frameshift, nonsense mutations, or deletions, compared to coding
246 regions that did not have these putatively deleterious mutations (functional) (Supplemental Figure 3) (all
247 pairwise comparisons P > 0.05; Mann-Whitney U test). Overall, we found no evidence of *cis*-regulatory
248 divergence associated with deleterious coding sequence divergence, suggesting *cis*-regulatory evolution
249 can evolve independently of Y-linked coding variation**.**

250 **Discussion**

251 The evolution of gene regulation on Y chromosomes has largely been studied in the context of
252 RNA expression levels of gametologs (Muyle, et al. 2012; White, et al. 2015; Beaudry, et al. 2017;
253 Muyle, et al. 2018; Martin, et al. 2019; Veltsos, et al. 2019). Over time, gametolog expression is generally
254 lost across most genes on the Y chromosome. With the growing number of Y chromosome reference
255 assembles and sequencing methods to quickly profile functional *cis*-regulatory regions across the genome,
256 it has become feasible to explore the molecular evolution of *cis*-regulatory regions that have led to altered
257 expression patterns. Here, we observed elevated nucleotide substitutions within *cis*-regulatory elements
258 across the threespine stickleback Y chromosome, consistent with positive selection driving this process.
259 We found that increased rates of *cis*-regulatory divergence were associated with X-biased expression of
260 gametologs, suggesting loss of expression from the Y chromosome. Without liver and testis gene
261 expression from the ninespine stickleback as an ancestral comparison, we were unable to confirm that X-
262 biased gametolog expression is a result of downregulation of the Y chromosome or upregulation of the X
263 chromosome. However, previous sequencing of a brain transcriptome showed expression patterns
264 consistent with loss of Y expression rather than gain of X expression (White, et al. 2015). It is therefore
265 likely that the X-biased expression we observed in liver and testis tissue also occurs through the loss of Y
266 chromosome expression, similar to what has been observed in other species with degenerating Y
267 chromosomes (Muyle, et al. 2012; Beaudry, et al. 2017; Wei and Bachtrog 2019). Additional work will be
268 necessary to identify the functional variants responsible for the X-biased expression pattern. We currently
269 cannot determine whether the X-biased expression pattern is due to the accumulation of deleterious
270 mutations from inefficient selection, adaptive substitutions that are accumulating in the ACRs, or a
271 combination of the two.

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

288 Evidence for adaptive silencing (Orr and Kim 1998) would be present if alleles with elevated
289 rates of coding divergence between the X and Y chromosomes, were more likely silenced on the Y
290 chromosome. Our results did not support this model. We found that the accumulation of X or Y ACR
291 divergence had no association with rates of coding evolution in nearby gametologs. Our results
292 complement previous findings that have compared coding sequence evolution to gametolog expression.
293 The level of expression from the Y chromosome is often not correlated with the overall number of
294 deleterious mutations within coding regions (Bachtrog 2006; Bachtrog, et al. 2008; Beaudry, et al. 2017).
295 We found *cis*-regulatory evolution was correlated with expression differences, but that ACR divergence
296 accumulated independent of coding divergence. We also found no correlation between divergence within
297 an ACR and the functional state of a coding region (i.e. containing frameshift or nonsense mutations). A
298 correlation between nucleotide substitutions in *cis*-regulatory regions and deleterious mutations in coding
299 regions may be observed if ACRs and gene expression are profiled from additional tissues. In addition,
300 we assigned ACRs to coding sequences based on overall proximity to a gene. Although this method of
301 annotating ACRs is commonly used to assign *cis*-regulatory regions to genes (Connelly, et al. 2014;
302 Alexandre, et al. 2018; Ricci, et al. 2019), it is possible that some ACRs interact with other genes through
303 long range interactions based on chromatin configuration (Mifsud, et al. 2015; Schoenfelder and Fraser
304 2019), reducing our ability to detect an association between deleterious mutations in coding regions and
305 nucleotide substitutions in ACRs. Also, the number of substitutions within an ACR may not be a fully

306 accurate predictor of the overall ability to silence a given gametolog on the Y chromosome. If only a
307 small number of substitutions are needed to ablate transcription factor binding within an ACR, we would
308 not expect to see a strong correlation between the number of deleterious mutations within coding
309 sequences and the number of substitutions within ACRs. Additional functional work will be necessary to
310 explore these alternatives.

311 The degeneration by regulatory evolution model would be supported if *cis*-regulatory regions
312 exhibited elevated substitution rates quickly following recombination suppression (Lenormand, et al.
313 2020; Lenormand and Roze 2022). Silencing of Y-linked alleles occurs first in this model, followed by
314 the accumulation of deleterious mutations within coding regions. We found that putative *cis*-regulatory
315 elements exhibited signatures of positive selection on the Y chromosome across all three evolutionary
316 strata, including the two youngest strata that still contain most of the ancestral gene content (Peichel, et al.
317 2020). The degeneration by regulatory evolution model predicts that the substitution rate within *cis*-
318 regulatory elements will accelerate once deleterious mutations begin to accumulate within coding regions.
319 This will further suppress gametolog expression from the Y chromosome, masking maladaptive coding
320 mutations. Consistent with this model, our results show *cis*-regulatory regions accumulate substitutions at
321 a faster rate than neutrally evolving regions of the chromosome. This suggests there is selection to
322 downregulate Y chromosome alleles.

323 The degeneration by regulatory evolution model also indicates dosage compensation should
324 evolve simultaneously for genes that are under strong stabilizing selection to maintain expression levels
325 (Lenormand, et al. 2020). Degeneration of *cis*-regulatory elements on the Y chromosome and loss of
326 expression should select for upregulation of the gametolog on the X chromosome to compensate for
327 dosage loss. In this scenario, nucleotide substitutions should accumulate within ACRs on the X
328 chromosome. We found that X-linked ACRs had an excess of substitutions compared to polymorphism,
329 consistent with positive selection. Unlike Y-linked ACRs, the X-linked ACRs had low divergence over
330 longer evolutionary timespans between ninespine and threespine stickleback fish. This suggests that
331 selection has acted on some X-linked sites more recently within threespine stickleback populations to fix
332 beneficial regulatory mutations. This distinction is consistent with Y-linked ACR divergence
333 accumulating first, as predicted by the degeneration by regulatory evolution model (Lenormand et al.
334 2020). Selection on X-linked *cis*-regulatory elements within ACRs may still be ongoing. A similar
335 finding was observed on the neo-sex chromosomes of *Drosophila miranda*, where the neo-X
336 chromosome was enriched for signatures of selective sweeps, while the ancestral X chromosome region
337 was not (Bachtrog, et al. 2009).

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

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

362 Expression balance of dosage sensitive genes can also be maintained if the Y-linked gametolog
363 does not degenerate. The coding sequence of haploinsufficient genes on Y chromosomes have been
364 shown to be maintained through purifying selection. These genes presumably cannot be lost from the Y
365 chromosome (Bellott, et al. 2014; Bellott, et al. 2017; Peichel, et al. 2020; Bellott and Page 2021). For
366 these genes, if Y-linked *cis*-regulatory elements accumulated deleterious mutations, ancestral expression
367 patterns could be maintained if compensatory transcription binding sites evolved. This would also result
368 in signatures of increased substitution rates with ACRs on the Y chromosome, relative to intergenic
369 regions. This type of compensatory evolution has been proposed for functionally critical genes in
370 mammals (Chaix, et al. 2008; Vermunt, et al. 2016), and *Drosophila* (Landry, et al. 2005; Arnold, et al.

371 2014; Signor and Nuzhdin 2018). Functional analysis of what transcription binding sites are affected by
372 Y-linked mutations could help identify which mutations are leading to down-regulation of gametologs
373 compared to those that maintain ancestral expression.

374 Recent models indicate inversions that suppress recombination between sex chromosomes can be
375 selected to be retained within populations if sex-specific *trans*-acting regulators evolve to maintain
376 optimal expression in both sexes (i.e. dosage compensation). (Lenormand and Roze 2022). This is due to
377 the fact that if divergent X- and Y- linked *cis*-regulators were to recombine, it would lead to maladaptive
378 expression levels in males or females. The evolutionary strata on the threespine stickleback Y
379 chromosome evolved from at least three nested inversions that suppressed recombination (Peichel, et al.
380 2020). Unlike this recent model, our results suggest these inversions were likely not under strong
381 selection to be retained initially because of early evolution of dosage compensation. Chromosome-wide
382 dosage compensation does not occur in the threespine stickleback (White, et al. 2015) and we did not
383 observe a corresponding evolution of X-linked upregulation. Interestingly, the excess of X-linked
384 mutations observed in testis ACRs, may be reflective of a greater abundance of rapidly evolving sex-
385 specific *trans*-regulators in the testes. However, additional work will be needed to identify the role that
386 autosomal *trans*-regulators play in modifying gene expression of sex chromosomes in each sex, to test
387 other aspects of the degeneration by regulatory evolution model.

388 Although we focused on the degeneration of *cis*-regulatory elements on the Y chromosome, it is
389 important to note that Y chromosomes become masculinized over time and *cis*-regulatory elements may
390 also be under positive selection for male-specific neo- or sub-functionalization (Soh, et al. 2014; Mahajan
391 and Bachtrog 2017; Bachtrog, et al. 2019; Martínez-Pacheco, et al. 2020; Peichel, et al. 2020; Chang, et
392 al. 2022). Some of the ACRs we found with high divergence may actually reflect functional *cis*-
393 regulatory regions associated with testis-specific genes. Indeed, some of the genes we observed with high
394 ACR divergence on the Y chromosome had a Y-biased expression pattern. Additional work will be
395 necessary to identify the function of these regulatory regions.

396 Y chromosomes often have much lower diversity overall, relative to neutral expectations
397 (Lawson Handley, et al. 2006; Wilson Sayres, et al. 2014). This pattern has been attributed to purifying
398 selection removing deleterious mutations and the linked neutral variation throughout the non-recombining
399 region (Wilson Sayres, et al. 2014; Wilson Sayres 2018). Our findings revealed that nucleotide diversity
400 on the threespine stickleback Y chromosome is also much lower than neutral expectations. Our results
401 suggest positive selection within ACRs could also be an important driver of reducing nucleotide diversity
402 on the Y chromosome within populations. One important consideration is whether sex-biased
403 demography may affect Y-linked diversity. If males are more variable in reproductive success than

404 females, this could lower the expected effective population size of the Y chromosome, reducing
405 nucleotide diversity. Little is known about the operating sex-ratios of stickleback fish. Some populations
406 exhibit sex ratios that are female biased (Rollins, et al. 2017). This could contribute to the low nucleotide
407 diversity we observed. However, simulations in human populations revealed that even drastic shifts in sex
408 ratio could not entirely explain the low diversity observed on the Y chromosome (Wilson Sayres, et al.
409 2014).

410 **Conclusion**

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

420

421 **Materials and Methods**

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

423 We used accessible chromatin regions (ACRs) from two different tissues. Liver ACRs were
424 previously identified from two replicates (Naftaly, et al. 2021) (NCBI BioProject PRJNA667175). We
425 also collected testes from two juvenile males (~4.4 cm in standard length) of laboratory-reared threespine
426 stickleback fish, originally isolated from Lake Washington (Seattle, Washington, USA) (NCBI BioProject
427 PRJNA686097). The testis cells were immediately dissociated through homogenization in 1X PBS
428 containing proteinase inhibitor cocktail (PIC, cOmplete tablets Roche) (PBS+PIC). The cells were fixed
429 with 16% formaldehyde and washed twice with PBS+PIC, followed by lysis in 1M Tris-HCl, pH=8, 0.5M
430 EDTA, 10% NP-40, 50% glycerol/molecular grade H_2O, and 1X PIC. Nuclei were stained with DAPI and
431 counted with a hemocytometer. We diluted samples to 60,000 – 80,000 nuclei. ATAC-seq library
432 preparation was conducted using previously established protocols (Lu, et al. 2017; Naftaly, et al. 2021).
433 The libraries were sequenced on an Illumina NextSeq (2 x 150 bp; Georgia Genomics and Bioinformatics
434 Core). Reads were trimmed with Trimmomatic (v. 0.36) (Bolger, et al. 2014), using a sliding window of

435 four bases, trimming the remainder of the read when the average quality within a window dropped below
436 20. Residual adapter sequences were removed using Trimmomatic ILLUMINACLIP. Reads were filtered
437 for a minimum length using MINLEN:30. We aligned the trimmed reads using Bowtie2 (v. 2.4.1) with
438 default parameters (Langmead and Salzberg 2012). We filtered for alignments with a mapping quality
439 greater than 20 using SAMtools (v1.14). We also removed PCR duplicates by using the MarkDuplicates
440 function in Picard (https://github.com/broadinstitute/picard).

441 We used a Tn5 control sample to normalize ATAC-seq reads to remove the effect of Tn5 bias.
442 For the control, genomic DNA was extracted from a caudal fin clip of one male fish using a standard
443 phenol-chloroform extraction. The ATAC-seq library preparation, whole genome sequencing library
444 preparation, and sequencing (Illumina HiSeq 2 x 150 bp) of the control sample were completed by
445 GENEWIZ (New Jersey, USA). Whole genome sequencing was conducted in order to show the DNA
446 sample was of sufficient quality to construct the Tn5 bias control. For the whole genome sequencing, we
447 recovered over 232 million reads with an average quality score of 38.59, indicating a high-quality sample.
448 We trimmed residual adapters and low-quality sequences using Trimmomatic as previously described.
449 Trimmed reads were aligned to the threespine stickleback genome using Bowtie2. The read coverage per
450 base pair was calculated using BEDTools (v2.29, -d) (Quinlan and Hall 2010). The whole genome
451 sequencing sample had an average read depth of 90x where only ~10% of the genome was supported with
452 less than ten reads. Within these regions, only 6% had zero reads per base pair, indicating 94% of the
453 genome could be queried for biased integration of Tn5. The Tn5 bias control produced over 177 million
454 reads. The reads were trimmed from these reads using Trimmomatic with the same parameters. The
455 trimmed reads were aligned to the threespine stickleback genome (v. 5) using Bowtie2 with default
456 parameters. Reads mapping to the mitochondria and unscaffolded regions were removed. PCR duplicates
457 were removed using MarkDuplicates from Picard.

458 We evaluated concordance between replicates in several ways. We first searched for enrichment
459 of ATAC-seq reads around the transcription start sites (TSSs) of expressed genes. Iso-Seq long read
460 sequencing was previously conducted in threespine stickleback fish to curate accurate transcription start
461 site annotations across multiple tissues (Naftaly, et al. 2021). We used deepTools (v3.5.1) computeMatrix
462 reference-point (Ramírez, et al. 2016) on the ATAC-seq alignments to assay read depth 3kb around the
463 complete set of annotated transcription start sites. We plotted enrichment of ATAC-seq reads around TSS
464 and compared to RNA-seq expression using deepTools plotHeatmap with default settings, except that we
465 used the -sortUsingsamples to sort the regions by expression. We found strong enrichment around TSSs,
466 across both replicates for the same set of expressed genes (Supplemental Figure 4). ACRs were called for
467 each replicate using MACS2 (v2.2.7.1) callpeak with the –keep--dup all parameter, and read depth was

468 normalized with Tn5 control sequencing with the -c parameter (Zhang, et al. 2008). The majority of
469 ACRs were shared across replicates (liver: 67% and testis: 61%) We then used the P-value of each
470 MACS2 peak to calculate the Irreproducibility Discovery Rate (IDR) (https://github.com/daniel-
471 shaw1/Regulatory_divergence_paper/blob/main/Supplemental/IDR.r), (Supplemental Figure 5). The IDR
472 determines the probability that a peak is reproducible between replicates, due to chance (Qunhua, et al.
473 2011; Landt, et al. 2012). Replicates that have IDR < 0.05 are generally considered to be highly
474 reproducible. However, the IDR test is considered very stringent (Jalili, et al. 2015), potentially removing
475 many true regions of increased accessibility. Because both replicates were exhibited high concordance,
476 we pooled the aligned reads from both replicates and called peaks from the joint alignment using MACS2
477 as previously described, as pooling raw reads has been previously reported (Reske, et al. 2020; Yan, et al.
478 2020) as a strategy to find high quality enriched regions. We found that most ACRs (liver: 79%, testis:
479 75%) from the pooled sample overlapped with low scoring ACRs from the individual replicates (IDR <
480 0.05). This suggested a majority of ACRs in the pooled dataset were of high quality in the individual
481 replicates.

482 To define an ancestral set of ACRs, we mapped the nucleotide sequence of each ACR on the X
483 chromosome to the homologous region on the Y chromosome using previously generated alignment
484 coordinates (Peichel, et al. 2020). A similar approach has been used previously to define the ancestral set
485 of coding regions, aligning X-linked gene annotations to the Y chromosome (Peichel, et al. 2020). We
486 also attempted to identify Y-linked ACRs and calculate divergence in a similar method as we identified
487 with X-linked ACRs. However, we found that MACS2 identified significantly fewer enriched regions of
488 accessible chromatin on the Y chromosome compared to the X chromosome. The few ACRs that were
489 identified also aligned to multiple regions on the X chromosome, suggesting these regions were repetitive.
490 Using the X chromosome ACRs, we identified orthologous regions between the X chromosome, Y
491 chromosome, and orthologous autosome (chromosome 19) from the ninespine stickleback fish *(Pungitius*
492 *pungitius*; (Varadharajan, et al. 2019) using BLAST+ (blastn v. 2.11.0) with default blastn parameters and
493 -perc_identity set to 75. (Camacho, et al. 2009). To ensure high quality alignments, we filtered for
494 uniquely mapping alignments that also had a bit score greater than 100. We used BLAST+ alignments to
495 extract sequence to create multiple sequence alignments for downstream analysis. We aligned the
496 threespine stickleback X and Y ACR sequences to the ninespine stickleback autosome sequence using
497 MUSCLE (v 3.8.1551) (Edgar 2004) with default parameters. We used MUSCLE to create a multiple
498 sequence alignment between the X sequence, Y sequence, and the sequence from the orthologous
499 autosome in order to identify X- and Y- specific variants. We called single nucleotide variants using snp-
500 sites (v2.5.1) (Page, et al. 2016) with default parameters. We calculated the reported divergence rate by
501 dividing the number of variants by the number of the aligned sites.

502 We compared ACR divergence on the sex chromosomes to ACR divergence on autosome 18, a
503 chromosome similar in length to the X chromosome. We identified orthologous regions between the
504 threespine stickleback autosome 18 and autosome 18 from the ninespine stickleback fish using BLAST+
505 (blastn v. 2.11.0) (Camacho, et al. 2009) as previously described. We called variants between autosomal
506 regions in a similar manner as the sex chromosomes, by creating a pairwise alignment using MUSCLE (v
507 3.8.1551) with default parameters and called single nucleotide variants using the same python script as
508 above. We calculated a divergence rate by dividing the number of single nucleotide variants by the size of
509 the aligned region.

510 Divergence within ACRs was compared with the divergence of coding regions of neighboring
511 genes. We identified the closest gene to each ACR by running annotatepeaks.pl in homer (v 4.11) (Heinz,
512 et al. 2010) using default settings. Per default settings, ACRs were assigned to the closest gene within
513 50kb of the TSS. We verified the homer peak annotations with a custom Python script that identifies
514 ACRs near the TSS of genes (https://github.com/daniel-
515 shaw1/Regulatory_divergence_paper/blob/main/Supplemental/Peaksneargenes.py). Both methods were
516 100% concordant with each other. We further verified a subset of the annotations (200 peaks total) by
517 visualizing the ACRs in the threespine stickleback JBrowse genome browser
518 (https://stickleback.genetics.uga.edu/). We found that 100% of the 200 X-linked ACRs were located
519 proximal to the predicted annotated gene (Supplemental Figure 6). For coding divergence comparisons,
520 we used previously reported estimates of synonymous (d_S) and non-synonymous (d_N) divergence of
521 gametologs between the X and Y chromosomes (Peichel et al., 2020). Coding regions with d_N/d_S ratios of
522 99 were omitted as these represent alignments with very little sequence divergence to estimate d_N/d_S
523 accurately.

524 **Estimating a control substitution rate**

525 To generate control regions, we randomly sampled non-functional intergenic regions throughout
526 the X chromosome, Y chromosome, or chromosome 18 to estimate a neutral substitution rate. Regulatory
527 regions tend to be GC-rich, which can be prone to higher mutation rates through mechanisms like GC-
528 biased gene conversion or spontaneous deamination of methylated cytosines (Nesta, et al. 2021). We
529 therefore GC-matched the randomly drawn intergenic regions with the GC content of ACRs. We
530 calculated the GC percentage of ACRs using a modified Perl script (countbp.pl, Nicholas Navin,
531 http://www.navinlab.com/bioperl/bioperl/gc_content.html). For each ACR, a random intergenic region
532 was drawn from the X chromosome equal in length and with a GC content within 2.0%. Intergenic
533 regions were defined as any region that fell outside of annotated functional regions from a combination of
534 Ensembl annotations (Cunningham, et al. 2021), Isoseq transcripts (Naftaly, et al. 2021), Y chromosome

535 annotations (Peichel, et al. 2020), repetitive elements (Peichel, et al. 2020; Nath, et al. 2021) and ACRs
536 from this study. This was performed for each stratum individually using bedtools shuffle –I ACRs.bed -
537 incl XYalignment.bed -excl annotations.bed. We extracted 10,000 sets of GC-matched intergenic regions
538 from each of the three evolutionary strata on the threespine stickleback sex chromosomes (set size same
539 as number of Liver ACRs: stratum one: 442, stratum two: 228, stratum three: 199; Testis ACRs: stratum
540 one: 30, stratum two: 37, stratum three: 51). Due to a limited number of intergenic sites shared between
541 the sex chromosomes, we sampled with replacement to align a sufficient number of intergenic regions.
542 We identified the orthologous regions to the ninespine genome assembly using BLAST+ (blastn v2.11.0)
543 (Camacho, et al. 2009). We filtered for alignments that mapped uniquely to chromosome 19 in the
544 ninespine stickleback fish, had an alignment length equal to 75% of the threespine query sequence, and
545 had a bit-score greater than 100. We generated multiple sequence alignments using MUSCLE (v
546 3.8.1551), for each intergenic permutation and calculated X and Y specific substitution rate as previously
547 described.

548 **Estimating within population nucleotide diversity**

549 We used whole-genome short-read sequencing of 12 males from the Lake Washington population
550 (Washington, USA; NCBI SRA SRP137809) to estimate nucleotide diversity across the sex
551 chromosomes. The raw reads were trimmed with Trimmomatic (v. 0.39) (Bolger, et al. 2014), using a
552 sliding window of four bases, trimming the remainder of the read when the average quality within a
553 window dropped below 15. The leading and trailing base pairs below quality three of every read were
554 removed along with any residual adapter sequence. After trimming, any reads below a minimum length of
555 36 were discarded. We aligned the trimmed reads using Bowtie2 (v. 2.4.5) with default parameters
556 (Langmead and Salzberg 2012). We also marked PCR duplicates by using the MarkDuplicates function in
557 Picard (v. 2.26.10) (https://github.com/broadinstitute/picard). SNP genotyping was conducted using the
558 GATK software package (v. 4.2.5.0). We called variants using HaplotypeCaller in GVCF mode. We then
559 joint-called variants using GenotypeGVCFs using a genomics database created by GenomicsDBImport.

560 We only considered biallelic SNPs in our estimates of nucleotide diversity and we filtered for
561 high quality genotypes using several different methods. On the sex chromosomes, sites should only have
562 hemizygous genotypes in males in the non-crossover region outside of the pseudoautosomal region. Any
563 sites that are heterozygous would be caused by errors in read alignment. We therefore did not consider
564 sites that were heterozygous in the non-crossover region on the X or Y chromosomes. We also filtered out
565 sites that exhibited too low or high of read depth, which would be indicative of alignment errors. To do
566 this, we did not consider sites that were less than one-half or greater than double the median read depth of
567 each chromosome (X and Y chromosomes: positions were retained if the read depth was between 3.5 and

568 14; chromosome 18: positions were retained if the read depth was between 6.5 and 26). Read filtering was
569 conducted using custom Perl scripts.

570 Nucleotide diversity (pi) was estimated on a per-site basis (--site-pi) across each chromosome
571 using vcftools (v. 0.1.16) (Danecek, et al. 2011)). Nucleotide diversity for ACRs was computed as the
572 average diversity across all sites within ACRs, including non-variant (zero pi) sites that were not removed
573 by read depth filters. We compared nucleotide diversity within ACRs with nucleotide diversity in
574 randomly drawn intergenic sites. Intergenic sites were defined as previously described. Intergenic sites
575 were drawn with replacement equal to the total number of sites within ACRs on each chromosome
576 (10,000 total permutations).

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

578 We performed a modified version of the McDonald-Kreitman test (Andolfatto 2005) by treating
579 divergence (D) and diversity (P) within ACRs as the non-synonymous factor, and non-functional
580 intergenic sequence previously generated was used as synonymous factor. We pooled sites from each
581 factor as previously described (Andolfatto 2005). We counted the total number of variants or fixed
582 substitutions (divergence) for each factor to calculate α, the fraction of substitutions that are likely fixed
583 due to positive selection. Divergence was estimated either between the threespine stickleback X and Y
584 chromosomes or between the threespine and ninespine stickleback autosome 18. Negative alpha values
585 were interpreted as regions that have rates of adaptive evolution near zero (Good, et al. 2013). We also
586 used these counts to estimate the neutrality index (NI) (Rand and Kann 1996). An NI greater than one
587 indicates an excess of ACR divergence within species, while an NI less than one indicates an excess of
588 ACR divergence between species (or between the X and Y). An index lower than one would be consistent
589 with positive selection. Standard deviation was determined through non-parametric bootstrapping. We
590 bootstrapped 10,000 replicates for each factor within the McDonald-Kreitman or NI test and computed
591 both metrics for each replicate.

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

593 We measured allele-specific expression of transcripts on the X and Y chromosome using three
594 replicate liver samples (NCBI BioProject PRJNA591630) and three replicate testis samples (NCBI
595 BioProject PRJNA591630). We trimmed the RNA-seq reads with Trimmomatic (v. 0.36) (Bolger, et al.
596 2014) using a sliding window of 4 bases, trimming the remainder of the read when the average quality
597 within a window dropped below 20 (SLIDINGWINDOW:4:20). Residual sequencing adapters were also
598 removed using ILLUMINACLIP. We aligned RNA-seq reads using Tophat2 (v2.2.1) with default
599 parameters (Kim, et al. 2013). We filtered for alignments with a map quality greater than 25 using

SAMtools (v1.14) (Li, et al. 2009) to identify reads that map uniquely on the X and Y allele. Similar map quality filters have been used to distinguish short reads between sex chromosomes in *Drosophila* and stickleback fish (Ellison and Bachtrog 2019; Peichel, et al. 2020) and between sexes with heteromorphic sex chromosomes in guppies (Kirkpatrick, et al. 2021). Read counts were obtained using htseq-count (v 0.9.1) (Anders, et al. 2015). Default parameters were used with the addition of --stranded=no and --nonunique all to maximize our ability to count reads on the Y chromosome. We used a custom GTF file for htseq-count that included previously determined start and end sites of all Ensembl predicted transcripts that are shared between the X and Y chromosomes (Peichel, et al. 2020). Reads were counted on the X and Y across the entire transcript as a feature. Transcripts were removed from the analysis if they had an RNA expression count of zero in all samples. For each tissue, we calculated the average read count for each gametolog across the three tissue replicates. We calculated the allele-specific expression as $\log_2$(X transcript average / Y transcript average).

Data availability

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

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

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 three anonymous reviewers for constructive feedback that improved the manuscript. We would also like to thank Bob Schmitz and Casey Bergman for helpful discussions related to the project.

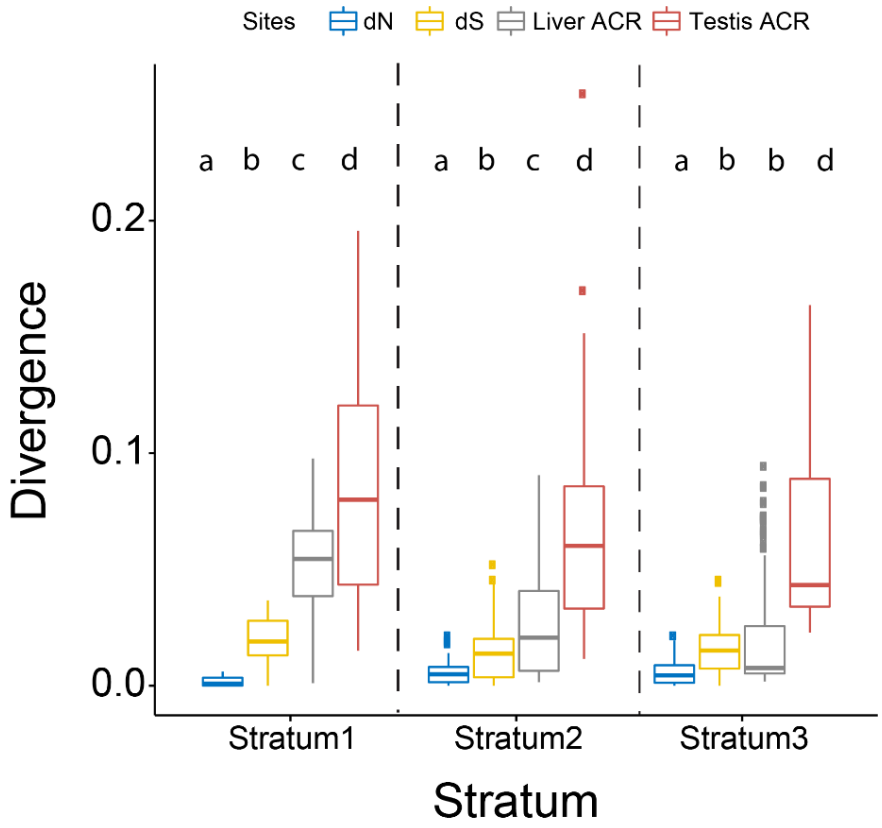
Declaration of interests

The authors declare no conflicts of interests.

625 **Figure legends**
626
627 **Figure 1. XY divergence within accessible chromatin regions and coding regions.** ACRs show
628 elevated divergence between the X and Y, relative to nonsynonymous and synonymous substitutions in
629 coding regions. Protein-coding divergence was estimated between the X and Y chromosomes at all shared
630 gametologs. Accessible Chromatin Regions (ACR) were identified on the X chromosome and mapped to
631 the Y chromosome. Substitutions within ACRs were summed and divided by the total length of the ACR
632 to calculate divergence. Stratum one: 12 gametologs, 442 liver ACRs, 30 testis ACRs; Stratum two: 71
633 gametologs, 228 liver ACRs, 37 testis ACRs; Stratum three: 46 gametologs, 199 liver ACRs 51 testis
634 ACRs. Rates of sequence divergence for all 4 categories were compared using a Kruskal-Wallis test for
635 each stratum ($P < 0.001$ for all strata). A post-hoc Dunn's test was used to identify which groups were
636 significantly different from one another. P- values were adjusted for multiple comparisons with a
637 Bonferroni correction. The letters above each plot indicate the significant differences between sites within
638 each stratum ($P < 0.001$). Box plots that share a letter assignment are not significantly different from one
639 another ($P > 0.001$).
640
641 **Figure 2 Accessible chromatin divergence between threespine and ninespine stickleback. A.**
642 Predicted substitution rates within accessible regions**.** ACR substitution rates (red) lower than random
643 non-functional control regions (white) and synonymous substitutions (blue) would indicate purifying
644 selection (left). ACR substitution rates roughly equal to non-functional control regions would indicate
645 mutations are accumulating due to inefficient selection from the loss of recombination (middle). ACR
646 substitution rates higher than non-functional control regions would indicate positive selection. **B.** Y-
647 linked ACR divergence is elevated over non-functional intergenic regions. X and Y ACR divergence was
648 compared to randomly drawn non-functional regions matched for GC content (10,000 permutations). The
649 average ACR divergence of each stratum is shown by the dotted line (stratum one: pink, stratum two:
650 green, and stratum three: blue). ACRs were identified on the X chromosome and mapped to the Y
651 chromosome and homologous ninespine stickleback autosome 19 from liver (top) and testis (bottom).
652 Autosome divergence is from chromosome 18. X- and Y- specific within ACRs were summed and
653 divided by the total length of the ACR to calculate divergence. Stratum one: 442 liver ACRs and 30 testis
654 ACRs; stratum two: 228 liver ACRs and 37 testis ACRs; stratum three: 129 liver ACRs and 46 testis
655 ACRs; autosomal: 1056 liver ACRs and 305 testis ACRs. Graphic made with BioRender.com
656
657 **Figure 3. Quantifying adaptive divergence and departure from neutrality among accessible**
658 **chromatin regions.** Sex-linked ACRs exhibited signatures of positive selection in both liver (left) and
659 testis (right). The neutrality index (NI) was calculated using the odds ratio of the McDonald Kreitman
660 test. $NI > 1$ reflects purifying selection (blue) while $NI < 1$ is a signature of positive selection (red). Error
661 bars represent two standard deviations calculated from 10,000 bootstrap replicates.
662
663 **Figure 4**. **Gametolog expression level compared to accessible chromatin region divergence.** ACR
664 divergence on the X and Y chromosomes is correlated with X-biased gene expression in both liver and
665 testis tissues. RNA-seq transcript counts that uniquely mapped to the X and Y chromosome were
666 quantified to determine allele-specific expression for all gametologs. Gametolog expression was
667 compared to the average divergence of Y- specific (A and B) and X-specific (C and D) mutations for
668 ACRs within 50 kb of expressed genes. (A). Y liver: $R= 0.40$, $P = 0.004$, $N = 65$; (B.) Y Testis: $R= 0.34$,
669 $P = 0.008$, $N = 65$; (C). X Liver: $R = 0.30$, $P = 0.033$, $N = 77$; (D). X Testis: $R = 0.25$, 0.061, $N = 77$. The
670 regressions were not significantly different between the X and Y chromosome for liver ($P = 0.590$, Fisher
671 transformation) or testis ($P = 0.077$, Fisher transformation).
672
673
674
675

676 **Figures**

677



678

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

680

681

682

683

684

685

686

687

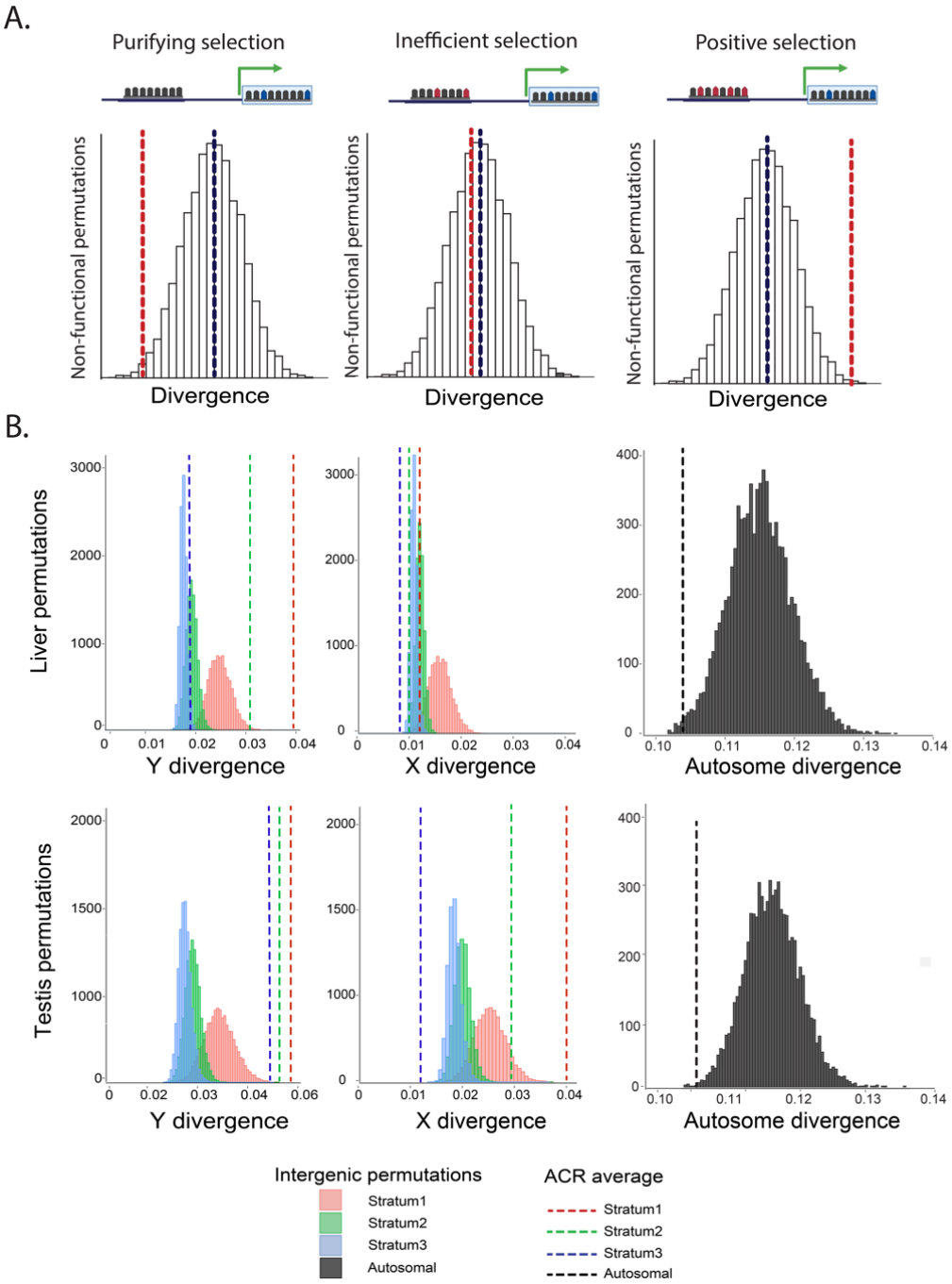
688

689

690

691

692



693

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

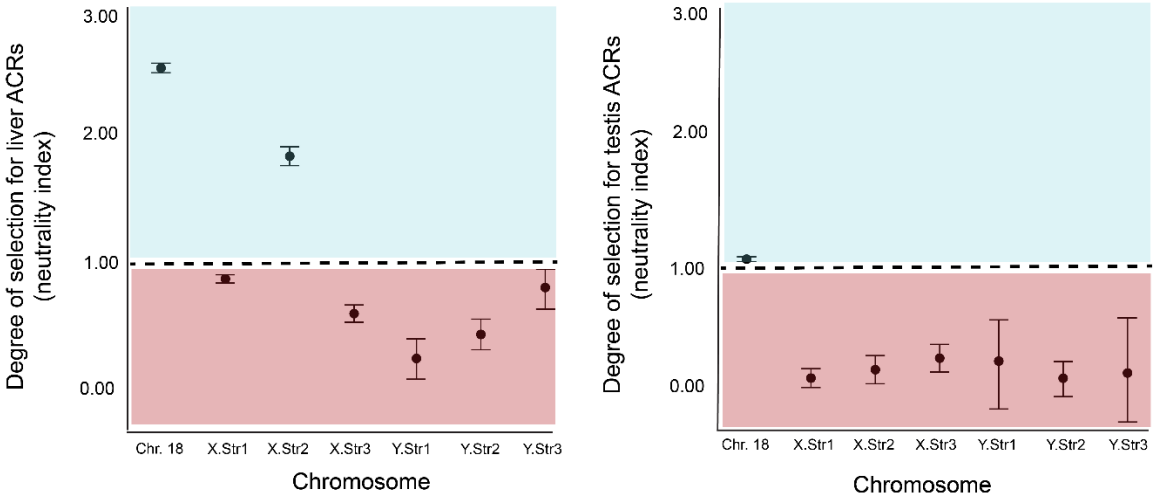


Figure 3. Departure from neutrality at sex-linked ACRs.

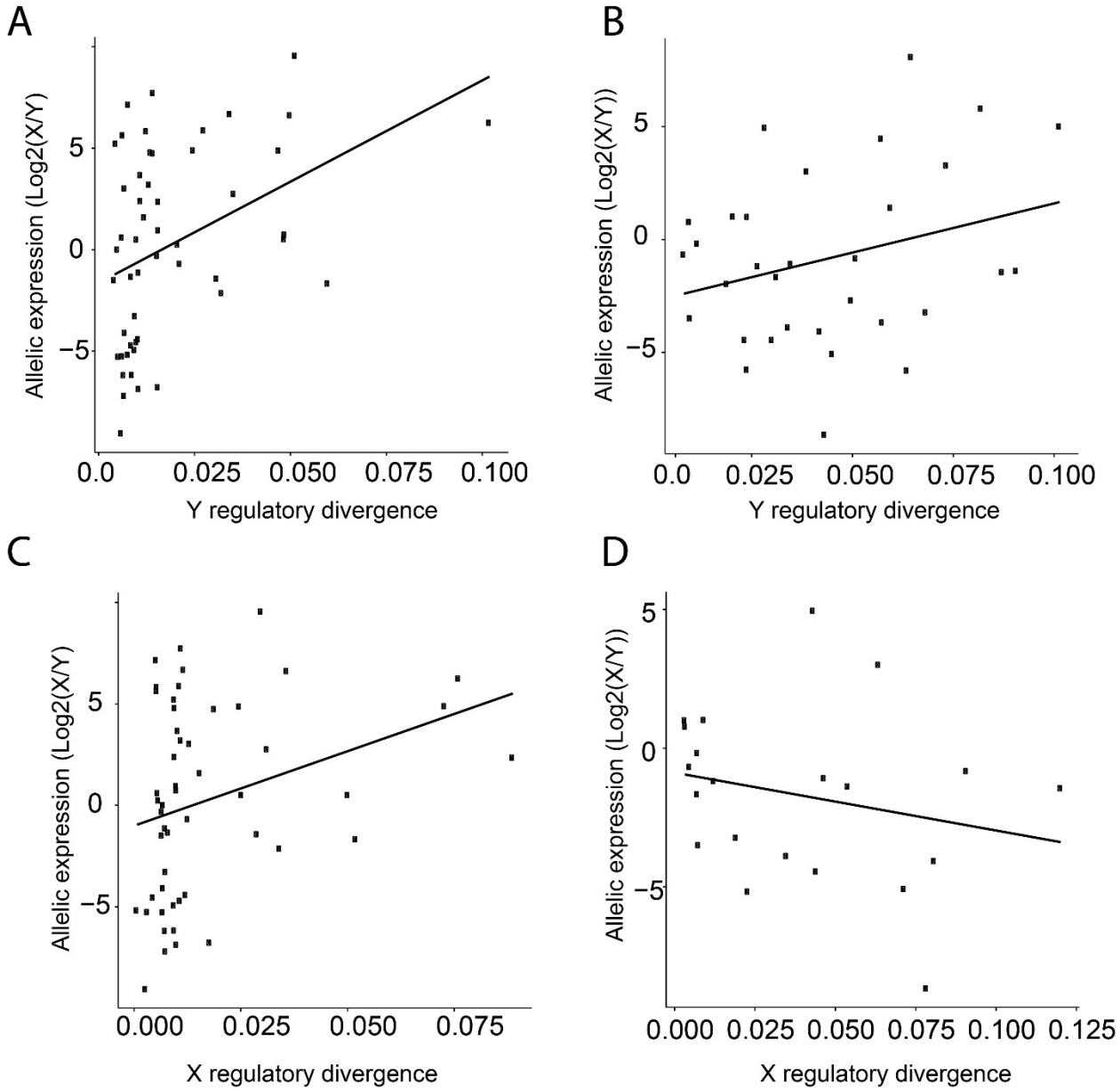


Figure 4. **Gametolog expression level compared to ACR divergence.**

703 **Table 1. Non-coding McDonald-Kreitman Test**

Chromosome	Alpha Liver ACRs	Neutrality Index Liver ACRs	Alpha Testis ACRs	Neutrality Index Testis ACRs	Standard Deviation (Liver, Testis)
Chr. Y Str. 1	0.845	0.155	0.710	0.290	0.080, 0.395
Chr. Y Str. 2	0.634	0.366	0.860	0.140	0.046, 0.140
Chr. Y Str. 3	0.222	0.778	0.815	0.185	0.127, 0.560
Chr. X Str. 1	0.214	0.786	0.978	0.022	0.040, 0.082
Chr. X Str. 2	-0.862	1.862	0.904	0.096	0.083, 0.123
Chr. X Str. 3	0.519	0.481	0.803	0.197	0.093, 0.121
Chr. 18	-1.637	2.637	-0.066	1.066	0.021, 0.021

704

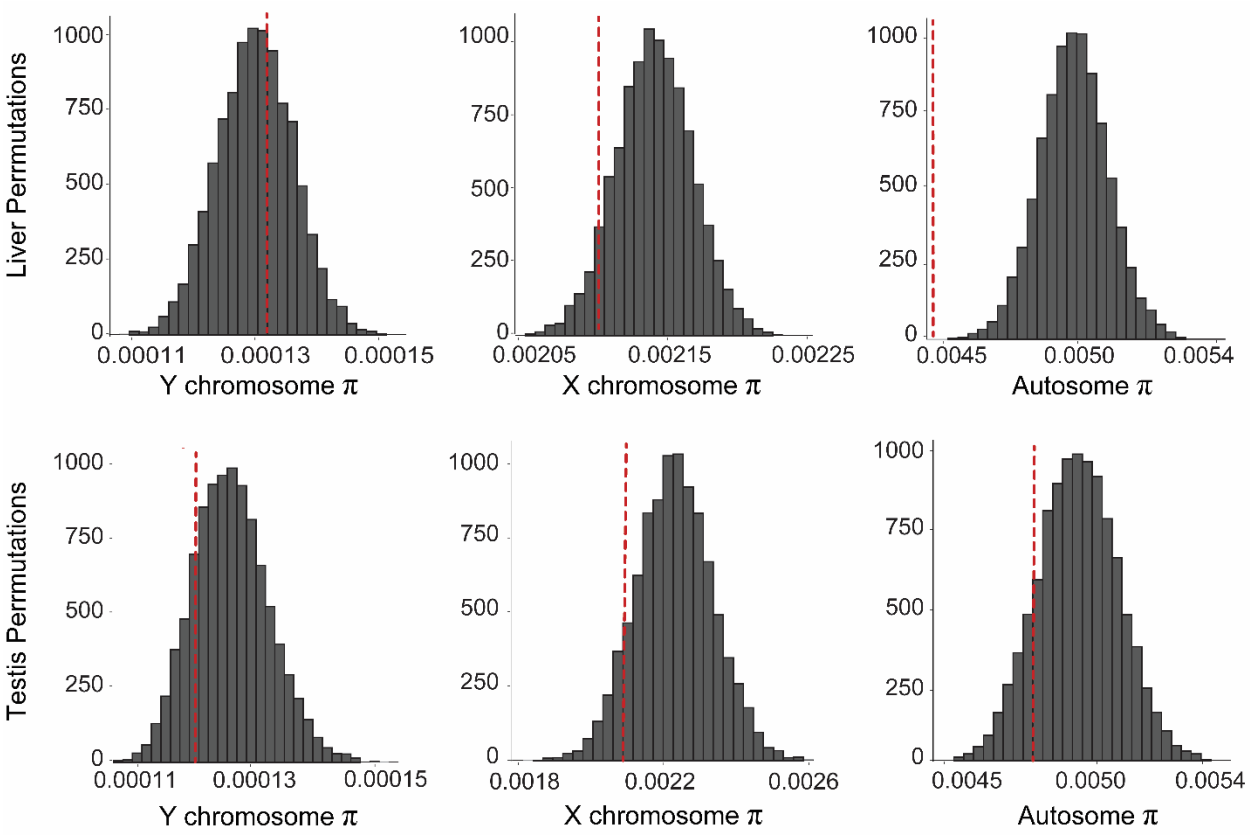
705 **Supplemental Figures**

706 **Supplemental Figure 1. Accessible chromatin region nucleotide diversity (pi) compared to**
707 **intergenic regions.** Nucleotide diversity within ACRs on the X and Y chromosome are not different than
708 randomly drawn non-functional regions (10,000 permutations). ACRs were identified on the X
709 chromosome and mapped to the Y chromosome in liver (top) and testis (bottom) Nucleotide diversity of
710 ACRs is also shown on chromosome 18. Average ACR nucleotide diversity for each chromosome is
711 shown by the dotted lines. Y chromosome ACR nucleotide diversity is not different from intergenic
712 permutations in liver (P = 0.623) or testis (P = 0.374). X chromosome ACR nucleotide diversity is not
713 different than intergenic regions in liver (P = 0.1985) or testis (P = 0.147). Autosomal ACRs exhibited
714 significantly lower diversity compared to intergenic permutations from liver, not from testis (Liver: P <
715 0.001, Testis P = 0.252). Stratum one: 442 liver ACRs and 30 testis ACRs; stratum two: 228 liver ACRs
716 and 37 testis ACRs; stratum three: 129 liver ACRs and 46 testis ACRs; autosomal: 1056 liver ACRs and
717 305 testis ACRs.
718
719 **Supplemental Figure 2. d_N / d_S ratios of coding regions compared to accessible chromatin region**
720 **divergence.** There is no correlation between d_N / d_S and ACR divergence using a Spearman's Rank
721 Correlation using ACRs from liver (A and B) or testis (C and D) tissue. Liver: (**A**) Y divergence:
722 (Stratum one (Blue): R=-0.12 P = 0.230, N = 30, Stratum two (Yellow): R = 0.2, P = 0.430, N=61;
723 Stratum three (Grey): R = 0.11, P = 0.770 N=126. (**B**) X divergence: (Stratum one (Blue): R = -0.080, P =
724 0.420, N=30; Stratum two (Yellow): R = 0.032, P = 0.9, N = 61; Stratum three (Grey): R = -0.054, P =
725 0.890, N = 126. Similar results in testis: (**C**) Y divergence: (Stratum one (Blue): R = -0.020, P = 0.650,
726 N=14; Stratum two (Yellow): R = -0.430, P = 0.160, N = 12; Stratum three (Grey): R = -0.140, P = 0.480,
727 N = 17. (**D**) X divergence: (Stratum one (Blue): R = 0.520, P = 0.120, N=14; Stratum two (Yellow): R = -
728 0.360, P = 0.210, N = 12; Stratum three (Grey): R = -0.020, P = 0.670, N = 17.
729
730 **Supplemental Figure 3. Accessible chromatin region divergence compared to functional states of Y-**
731 **linked gametologs.** ACR divergence was not significantly different among gametologs that contained
732 mutations that potentially impede function (i.e., premature stop codons, frameshifts, deleted nucleotides)
733 or are functional. Y-specific functional mutations were identified by comparing coding sequence between
734 the Y chromosome and the outgroup autosomal coding sequence in the ninespine stickleback. ACRs were
735 assigned to the closest gametolog within 50kb. P > 0.05 for all pairwise comparisons for both liver (A and
736 C) and testis (B and D) for regulatory divergence on the X and Y.
737

Supplemental Figure 4. High concordance of chromatin accessibility around transcription start site of expressed genes. Transcription start sites (TSSs) for all genes were ordered by RNA-seq read depth. ATAC-seq read depth was quantified ±3 kb of TSSs. Similar ATAC-seq read depth was observed between replicates. ATAC-seq depth across replicates is a strong predictor of expression from RNA-seq from same tissue type.

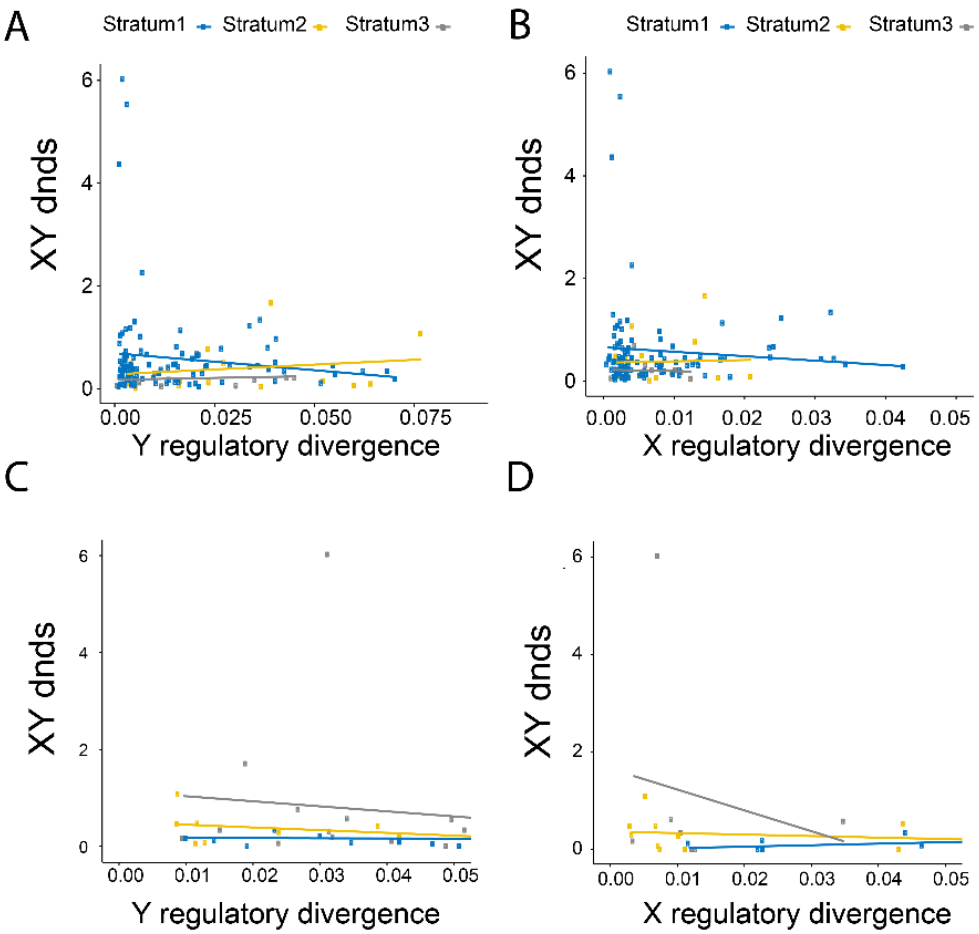
Supplemental Figure 5. High concordance of chromatin accessibility between peaks identified in two replicates. Peaks were identified in two replicates of liver (left) and testis (right). Log_{10}(P-Value) of each shared peak is highly correlated between replicates. Irreproducible Discovery Rate was calculated by calculated the ranked significance of reproducible and irreproducible peaks between replicates. Peaks with an IDR < 0.05 are shown in red.

Supplemental Figure 6. Verification of ATAC-seq proximity to annotated genes. Three representative examples of ATAC-seq read depth enriched at ACRs called with MACS2. The most proximal genes were identified with homer and with a custom python script. Annotations for a subset of 100 X chromosome ACR were visually inspected with JBrowse, to confirm that the peak was accurately assigned to nearby genes. ATAC-seq read depth is shown in grey. Bed files for the summit of each ACR identified by MACS2 is shown below the Ensembl gene annotations.



Supplemental Figure 1. Regulatory nucleotide diversity (pi) compared to intergenic regions

759

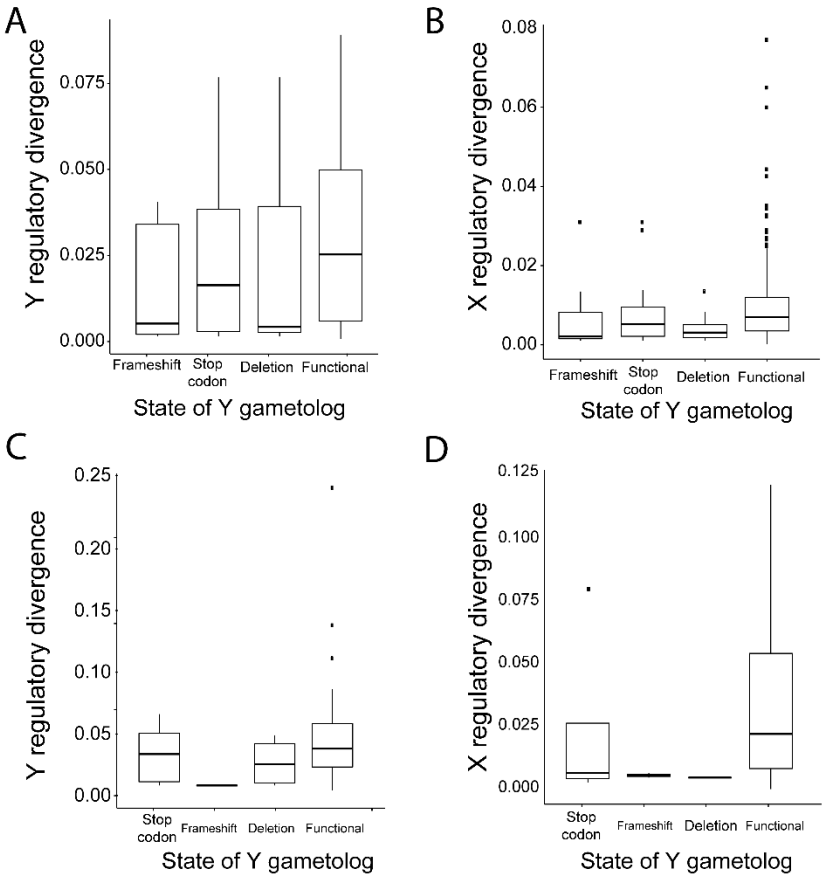


760

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

762

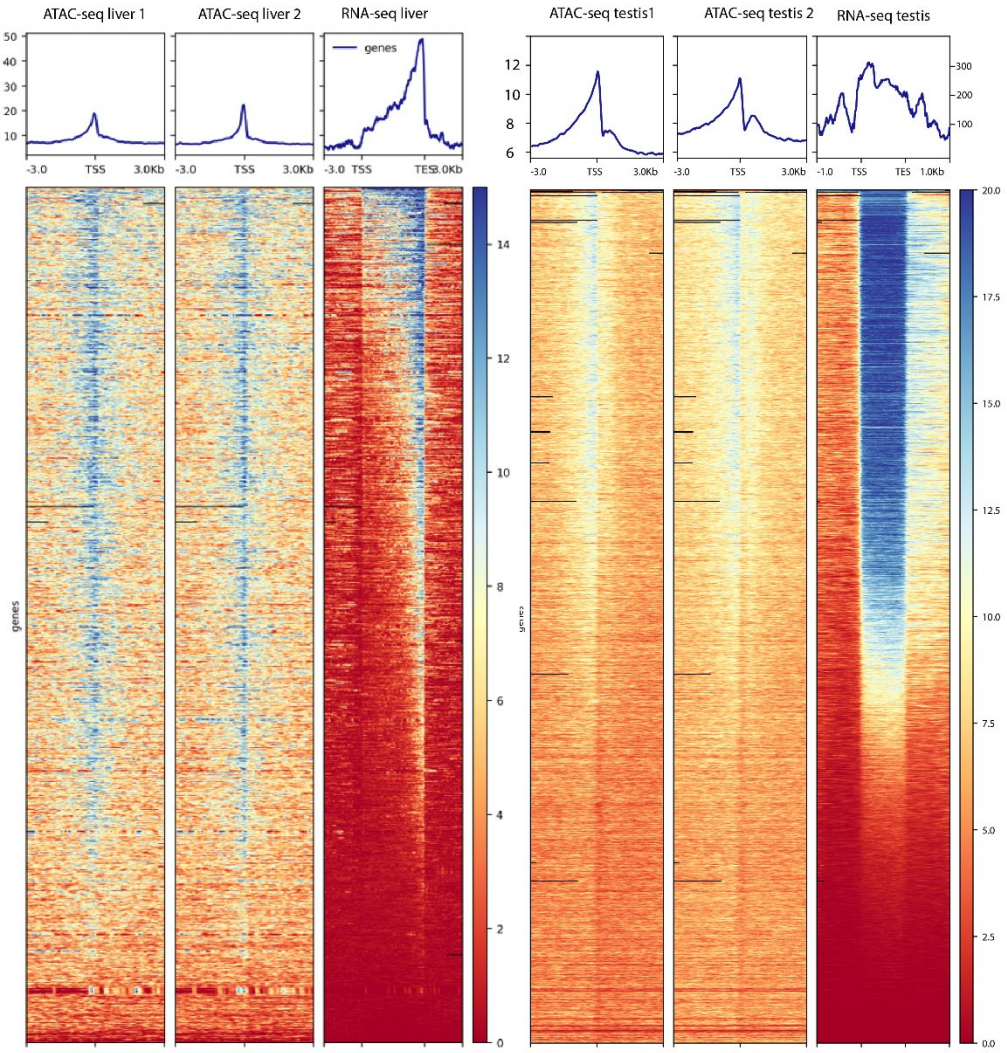
763



764

765 **Supplemental Figure 3. ACR divergence compared to functional states of Y-linked gametologs.**

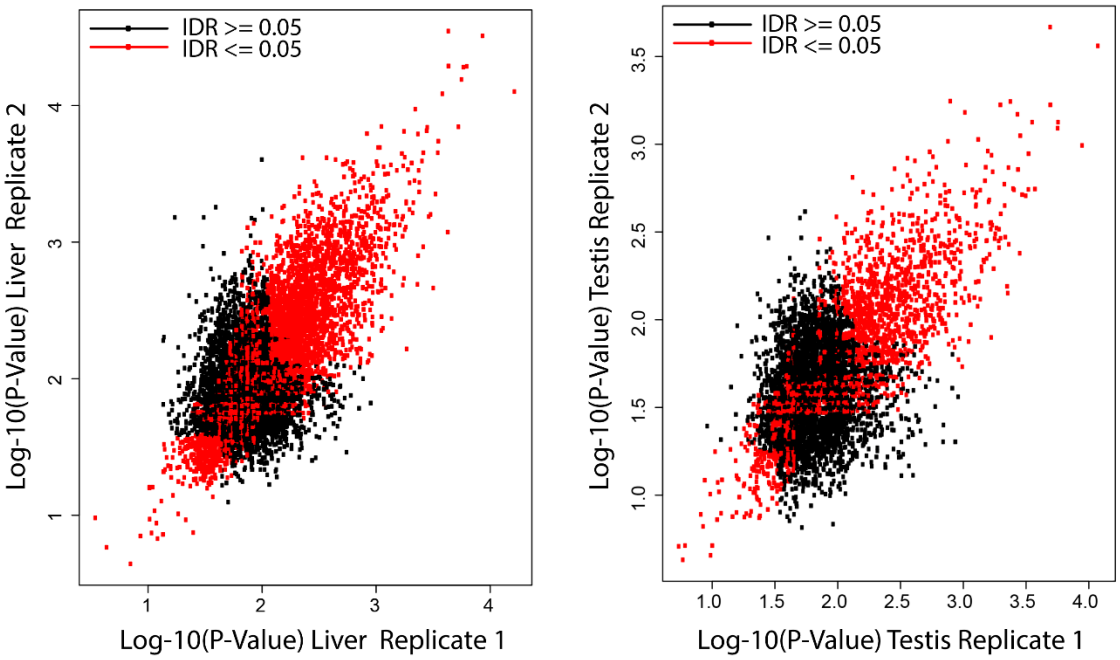
766



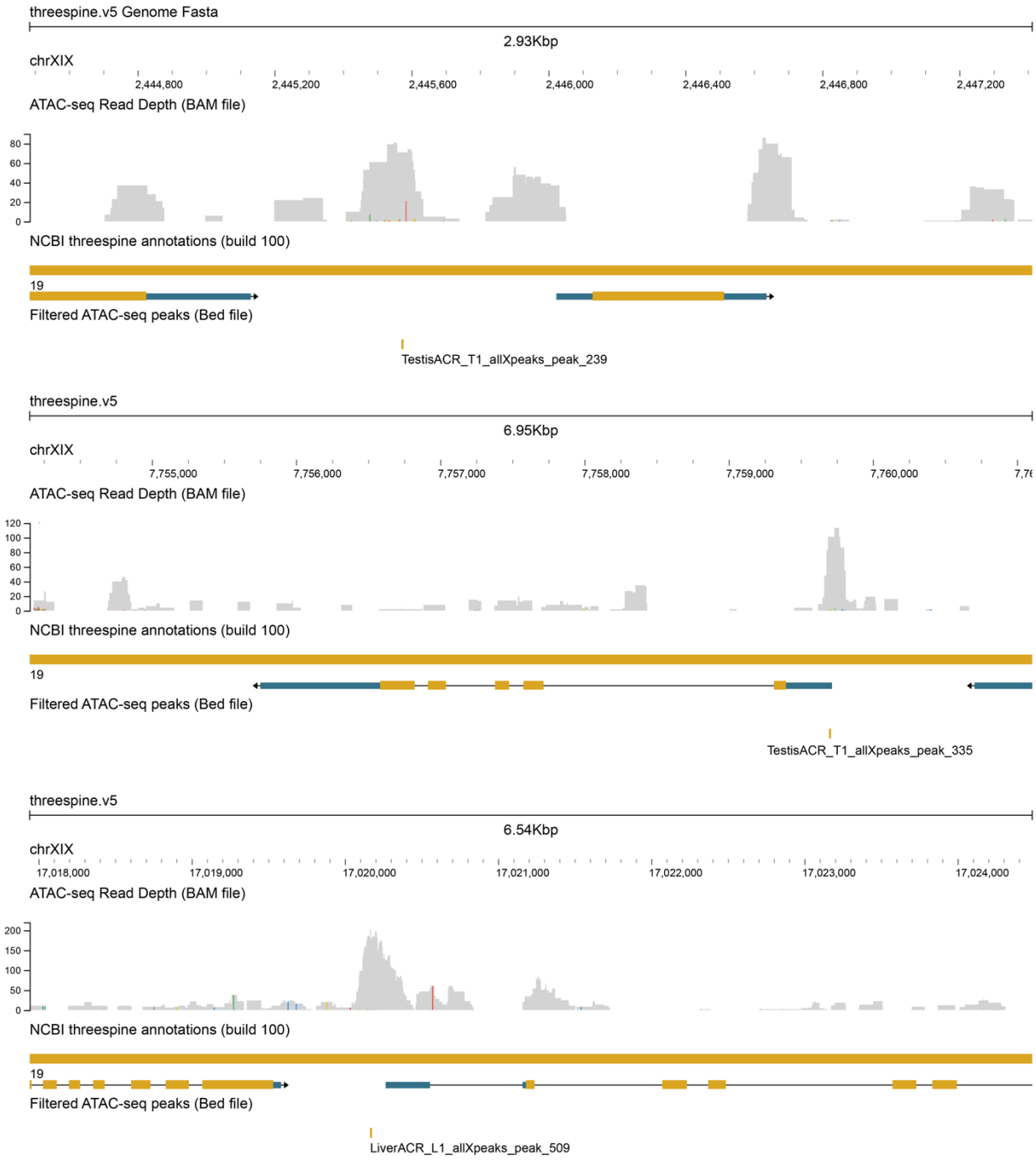
767

768 **Supplemental Figure 4. High concordance of chromatin accessibility around transcription start site**
769 **of expressed genes.**

770



Supplemental Figure 5. High concordance of chromatin accessibility between peaks identified in two replicates.



Supplemental Figure 6. Verification of ATAC-seq proximity to annotated genes.

780
781 **References**
782

783 Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, Sullivan AM, Bemm F,
784 Jolic D, Arsovski AA, et al. 2018. Complex Relationships between Chromatin Accessibility,
785 Sequence Divergence, and Gene Expression in Arabidopsis thaliana. Molecular Biology and
786 Evolution 35:837-854.
787 Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing
788 data. Bioinformatics 31:166-169.
789 Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149-1152.
790 Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. 2014. Quantitative
791 genome-wide enhancer activity maps for five Drosophila species show functional enhancer
792 conservation and turnover during cis-regulatory evolution. Nature Genetics 46:685-692.
793 Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. 2013. RNA
794 sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken
795 and allows comprehensive annotation of the W-chromosome. Genome Biology 14:R26.
796 Bachtrog D. 2006. Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. Current
797 Biology 16:1694-1699.
798 Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in
799 Drosophila. Genome Research 15:1393-1401.
800 Bachtrog D. 2008. The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration.
801 Genetics 179:1513-1525.
802 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome
803 degeneration. Nature Reviews Genetics 14:113-124.
804 Bachtrog D, Hom E, Wong KM, Maside X, De Jong P. 2008. Genomic degradation of a young Y
805 chromosome in Drosophila miranda. Genome Biology 9:R30.
806 Bachtrog D, Jensen JD, Zhang Z. 2009. Accelerated Adaptive Evolution on a Newly Formed X
807 Chromosome. PLoS Biology 7:e1000082.
808 Bachtrog D, Mahajan S, Bracewell R. 2019. Massive gene amplification on a recently formed Drosophila
809 Y chromosome. Nature Ecology & Evolution 3:1587-1597.
810 Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I,
811 Ming R, et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLoS Biology
812 12:e1001899.
813 Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008. Effects of X-Linkage and Sex-Biased Gene Expression on
814 the Rate of Adaptive Protein Evolution in Drosophila. Molecular Biology and Evolution 25:1639-
815 1650.
816 Beaudry FEG, Barrett SCH, Wright SI. 2017. Genomic Loss and Silencing on the Y Chromosomes of
817 Rumex. Genome Biology and Evolution 9:3345-3355.
818 Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T,
819 Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive
820 regulators. Nature 508:494-499.
821 Bellott DW, Page DC. 2021. Dosage-sensitive functions in embryonic development drove the survival of
822 genes on sex-specific chromosomes in snakes, birds, and mammals. Genome Research 31:198-
823 210.
824 Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, Galkina S, Pyntikova T, Koutseva N, Graves T,
825 et al. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive
826 regulators. Nature Genetics 49:387-394.

827 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
828 Bioinformatics 30:2114-2120.
829 Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A,
830 Kircher M. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343-
831 348.
832 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for
833 fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and
834 nucleosome position. Nature Methods 10:1213-1218.
835 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
836 architecture and applications. BMC Bioinformatics 10:421.
837 Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA. 2008. Evolution of Primate Gene Expression: Drift
838 and Corrective Sweeps? Genetics 180:1379-1389.
839 Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. 2022. Unique structure and positive
840 selection promote the rapid divergence of Drosophila Y chromosomes. eLife 11.
841 Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. Proceedings of
842 the National Academy of Sciences 75:5618-5622.
843 Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philosophical Transactions
844 of the Royal Society of London. Series B: Biological Sciences 355:1563-1572.
845 Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and
846 autosomes. The American Naturalist 130:113-146.
847 Connelly CF, Wakefield J, Akey JM. 2014. Evolution and Genetic Architecture of Chromatin Accessibility
848 and Function in Yeast. PLoS Genetics 10:e1004427.
849 Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. 2015. Molecular mechanisms
850 and evolutionary processes contributing to accelerated divergence of gene expression on the
851 Drosophila X chromosome. Molecular Biology and Evolution 32:2605-2615.
852 Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. 2013. Analysis of expressed SNPs identifies
853 variable extents of expression from the human inactive X chromosome. Genome Biology
854 14:R122.
855 Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, Austine-Orimoloye O,
856 Azov Andrey G, Barnes I, Bennett R, et al. 2021. Ensembl 2022. Nucleic Acids Research 50:D988-
857 D995.
858 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT,
859 Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.
860 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
861 Acids Research 32:1792-1797.
862 Ellison C, Bachtrog D. 2019. Recurrent gene co-amplification on Drosophila X and Y chromosomes. PLoS
863 Genetics 15:e1008251.
864 Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R,
865 Fischer A, et al. 2013. Comparative Population Genomics of the Ejaculate in Humans and the
866 Great Apes. Molecular Biology and Evolution 30:964-976.
867 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010.
868 Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory
869 Elements Required for Macrophage and B Cell Identities. Molecular Cell 38:576-589.
870 Horvath R, Josephs EB, Pesquet E, Stinchcombe JR, Wright SI, Scofield D, Slotte T. 2021. Selection on
871 Accessible Chromatin Regions in Capsella grandiflora. Molecular Biology and Evolution 38:5563-
872 5575.

873 Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki
874 C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y
875 chromosomes. Nature 483:82-86.
876 Hughes JF, Skaletsky H, Pyntikova T, Graves TA, Van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke
877 DP, Friedman C, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent
878 in structure and gene content. Nature 463:536-539.
879 Hughes JF, Skaletsky H, Pyntikova T, Koutseva N, Raudsepp T, Brown LG, Bellott DW, Cho T-J, Dugan-
880 Rocha S, Khan Z, et al. 2020. Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms
881 races in mammalian lineages. Genome Research 30:1716-1726.
882 Jalili V, Matteucci M, Masseroli M, Morelli MJ. 2015. Using combined evidence from replicates to
883 evaluate ChIP-seq peaks. Bioinformatics 31:2761-2769.
884 Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S.
885 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and
886 chimpanzees. Science 309:1850-1854.
887 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of
888 transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology
889 14:R36.
890 Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Schartl M. 2021. Evolution of the canonical sex
891 chromosomes of the guppy and its relatives. G3 Genes|Genomes|Genetics.
892 Kopania EEK, Larson EL, Callahan C, Keeble S, Good JM. 2022. Molecular Evolution across Mouse
893 Spermatogenesis. Molecular Biology and Evolution 39.
894 Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory cis-trans
895 Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila.
896 Genetics 171:1813-1822.
897 Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB,
898 Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE
899 consortia. Genome Research 22:1813-1831.
900 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
901 Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BA, Smith AD, Dean MD, Good JM. 2016. Contrasting
902 Levels of Molecular Evolution on the Mouse X Chromosome. Genetics 203:1841-1857.
903 Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong Purifying Selection at Synonymous Sites in
904 D. melanogaster. PLoS Genetics 9:e1003527.
905 Lawson Handley LJ, Berset-BräNdli L, Perrin N. 2006. Disentangling Reasons for Low Y Chromosome
906 Variation in the Greater White-Toothed Shrew (Crocidura russula). Genetics 173:935-942.
907 Lenormand T, Fyon F, Sun E, Roze D. 2020. Sex Chromosome Degeneration by Regulatory Evolution.
908 Current Biology 30:3001-3006.e3005.
909 Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual
910 dimorphism. Science 375:663-666.
911 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The
912 Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
913 Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. 2017. Combining ATAC-seq with nuclei sorting
914 for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Res 45:e41.
915 Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. 2019. The prevalence, evolution and
916 chromatin signatures of plant regulatory elements. Nature Plants 5:1250-1259.
917 Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. Nature
918 Communications 8.

919 Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young Drosophila Y
920 chromosome using single-molecule sequencing and chromatin conformation capture. PLoS
921 Biology 16:e2006348.
922 Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GAB, Touzet P. 2019. Evolution of
923 Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination
924 Systems. Genome Biology and Evolution 11:350-361.
925 Martínez-Pacheco M, Tenorio M, Almonte L, Fajardo V, Godínez A, Fernández D, Cornejo-Páramo P,
926 Díaz-Barba K, Halbert J, Liechti A, et al. 2020. Expression Evolution of Ancestral XY Gametologs
927 across All Major Groups of Placental Mammals. Genome Biology and Evolution 12:2015-2028.
928 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends in Genetics
929 29:537-544.
930 Meisel RP, Malone JH, Clark AG. 2012a. Disentangling the relationship between sex-biased gene
931 expression and X-linkage. Genome Research 22:1255-1265.
932 Meisel RP, Malone JH, Clark AG. 2012b. Faster-X Evolution of Gene Expression in Drosophila. PLoS
933 Genetics 8:e1003013.
934 Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey
935 W, Ewels PA, et al. 2015. Mapping long-range promoter contacts in human cells with high-
936 resolution capture Hi-C. Nature Genetics 47:598-606.
937 Muller HJ. 1918. Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal
938 Factors. Genetics 3:422-499.
939 Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid De Novo Evolution of X
940 Chromosome Dosage Compensation in Silene latifolia, a Plant with Young Sex Chromosomes.
941 PLoS Biology 10:e1001308.
942 Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Hobza R, Picard F, Widmer A,
943 et al. 2018. Genomic imprinting mediates dosage compensation in a young plant XY system.
944 Nature Plants 4:677-680.
945 Naftaly AS, Pau S, White MA. 2021. Long-read RNA sequencing reveals widespread sex-specific
946 alternative splicing in threespine stickleback fish. Genome Research:gr.274282.274120.
947 Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using
948 long-read sequencing. G3 Genes|Genomes|Genetics 11.
949 Nesta AV, Tafur D, Beck CR. 2021. Hotspots of Human Mutation. Trends in Genetics 37:717-729.
950 Orr HA, Kim Y. 1998. An Adaptive Hypothesis for the Evolution of the Y Chromosome. Genetics
951 150:1693-1698.
952 Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient
953 extraction of SNPs from multi-FASTA alignments. Microbial Genomics 2.
954 Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression.
955 Nature Reviews Genetics 14:83-87.
956 Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM,
957 Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals
958 convergent signatures of sex chromosome evolution. Genome Biology 21.
959 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
960 Bioinformatics 26:841-842.
961 Qunhua L, James BB, Haiyan H, Peter JB. 2011. Measuring reproducibility of high-throughput
962 experiments. The Annals of Applied Statistics 5:1752-1779.
963 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016.
964 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids
965 Research 44:W160-W165.

Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Molecular Biology and Evolution 13:735-748.

Reske JJ, Wilson MR, Chandler RL. 2020. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. Epigenetics & Chromatin 13.

Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, et al. 2019. Widespread long-range cis-regulatory elements in the maize genome. Nature Plants 5:1237-1249.

Rodríguez Lorenzo JL, Hobza R, Vyskot B. 2018. DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant Silene latifolia. BMC Genomics 19.

Rollins JL, Chiang P, Waite JN, Von Hippel FA, Bell MA. 2017. Jacks and jills: Alternative life-history phenotypes and skewed sex ratio in anadromous threespine stickleback (Gasterosteus aculeatus). Evolutionary Ecology Research 18:363-382.

Sardell JM, Josephson MP, Dalziel AC, Peichel CL, Kirkpatrick M. 2021. Heterogeneous Histories of Recombination Suppression on Stickleback Sex Chromosomes. Molecular Biology and Evolution 38:4403-4418.

Schoenfelder S, Fraser P. 2019. Long-range enhancer–promoter contacts in gene expression control. Nature Reviews Genetics 20:437-455.

Shaw DE, White MA. 2022. The evolution of gene regulation on sex chromosomes. Trends in Genetics.

Signor SA, Nuzhdin SV. 2018. The Evolution of Gene Expression in cis and trans. Trends in Genetics 34:532-544.

Singh ND, Koerich LB, Carvalho AB, Clark AG. 2014. Positive and Purifying Selection on the Drosophila Y Chromosome. Molecular Biology and Evolution 31:2612-2623.

Soh YQ, Alföldi J, Pyntikova T, Laura, Graves T, Patrick, Robert, Kremitzki C, Koutseva N, Jacob, et al. 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. Cell 159:800-813.

Tomaszkiewicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D, O'Brien PCM, Chikhi R, Ryder OA, et al. 2016. A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. Genome Research 26:530-540.

Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, et al. 2017. Landscape of X chromosome inactivation across human tissues. Nature 550:244-248.

Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS, Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. Genome Biology and Evolution.

Veltsos P, Ridout KE, Toups MA, González-Martínez SC, Muyle A, Emery O, Rastas P, Hudzieczek V, Hobza R, Vyskot B, et al. 2019. Early Sex-Chromosome Evolution in the Diploid Dioecious Plant Mercurialis annua. Genetics 212:815-835.

Vermunt MW, Tan SC, Castelijns B, Geeven G, Reinink P, De Bruijn E, Kondova I, Persengiev S, Bontrop R, Cuppen E, et al. 2016. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. Nature Neuroscience 19:494-503.

Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model. Evolution: International Journal of Organic Evolution 63:2413-2426.

Wei KHC, Bachtrog D. 2019. Ancestral male recombination in Drosophila albomicans produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. PLoS Genetics 15:e1008502.

White MA, Kitano J, Peichel CL. 2015. Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. Molecular Biology and Evolution 32:1981-1995.

1014 Wilson Sayres MA. 2018. Genetic Diversity on the Sex Chromosomes. Genome Biology and Evolution
1015 10:1064-1078.
1016 Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural Selection Reduced Diversity on Human Y
1017 Chromosomes. PLoS Genetics 10:e1004064.
1018 Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data
1019 analysis. Genome Biology 21.
1020 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li
1021 W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9:R137.

1022

1023

1024