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 across gametologs on heteromorphic sex chromosomes (X and Y or Z and W). Current models of sex chromosome evolution suggest this occurs through the rapid accumulation of mutations within regulatory regions. However, these patterns have not been demonstrated empirically due to the limited number of Y chromosome assemblies available to survey sequence evolution outside of coding regions. The threespine stickleback (*Gasterosteus aculeatus*) Y chromosome is an ideal model to test hypotheses of rapid regulatory evolution 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 (accessible chromatin regions defined through ATAC-seq) on the Y chromosome exhibited elevated substitution rates when compared to intergenic regions and synonymous sites within coding regions. This strongly suggests that many 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 regulatory regions exhibit signs of positive selection quickly after the suppression of recombination, supporting 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 across 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 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. 2012; 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 many deleterious regulatory mutations may accumulate through selective interference (Charlesworth and Charlesworth 2000; Bachtrog 2008), selection may also favor mutations within 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 empirically testing theoretical models that focus on regulatory evolution is 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 regulatory regions on the sex-limited chromosome. 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 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 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 functional elements for gene regulation (Connelly, et al. 2014; Lu, et al. 2019; Horvath, et al. 2021).

78 Profiling ACRs on sex chromosomes would therefore provide a means to study the molecular evolution
79 of *cis*-regulatory regions in the context of Y degeneration.

80 The threespine stickleback fish (*Gasterosteus aculeatus*) is an excellent model to study the
81 evolution of regulatory regions on sex chromosomes. Threespine stickleback fish have a high-quality
82 reference Y chromosome that is more recently derived compared to many other chromosome-scale
83 assemblies (Hughes, et al. 2010; Hughes, et al. 2012; Bellott, et al. 2014; Bellott, et al. 2017; Hughes, et
84 al. 2020; Peichel, et al. 2020). The Y chromosome contains multiple evolutionary strata at different stages
85 of degeneration (White, et al. 2015; Peichel, et al. 2020), allowing for comparisons of sequence evolution
86 at different temporal scales. While some of the Y chromosome has degenerated, over half of X-linked
87 gametologs remain on the Y chromosome (Peichel, et al. 2020). Many of the coding regions of the shared
88 gametologs show signatures of purifying selection and are enriched for dosage sensitive functions (White,
89 et al. 2015; Peichel, et al. 2020). Interestingly, chromosome-wide dosage compensation has not evolved
90 to counter the loss of expression for genes where the coding sequence has degenerated on the Y
91 chromosome (White, et al. 2015). The lack of chromosome-wide dosage compensation enables a gene by
92 gene comparison of *cis*-regulatory changes to understand how expression of X-linked alleles evolves in
93 response to degenerating coding regions on the Y.

94 Here, we leveraged ATAC-seq from two different tissues to characterize ACRs shared between
95 the X and Y chromosomes. To understand how selection may be operating within these regions, we
96 estimated nucleotide divergence of ACRs between the X and Y chromosomes of the threespine
97 stickleback fish (*Gasterosteus aculeatus*) as well as within-population nucleotide diversity of ACRs. We
98 found signatures of positive selection across sex-linked ACRs that indicate a rapid accumulation of
99 nucleotide substitutions on the Y chromosome, relative to other intergenic regions. Our findings
100 complement existing sex chromosome evolution theory that suggests the accumulation of mutations
101 within regulatory regions may be beneficial to silence expression from a degenerating Y chromosome.

102

103 **Results**

104 **Nucleotide divergence between the X and Y is higher in accessible chromatin regions**

105 The completion of the threespine stickleback Y chromosome assembly (Peichel, et al. 2020)
106 allowed us to thoroughly examine regulatory evolution among sex linked gametologs shared between the
107 sex chromosomes. Accessible chromatin regions (ACRs) are sites of the genome that putatively serve as
108 domains for transcription factor binding. We first sought to compare divergence of ACRs between the X

109 and Y chromosomes to previously characterized synonymous divergence throughout coding regions.
110 Crossing over was suppressed across the threespine stickleback sex chromosomes in at least three
111 separate events, forming strata with distinct levels of divergence (stratum one: oldest; stratum two:
112 middle; stratum three: youngest) (Peichel, et al. 2020). Based on theories of regulatory evolution on sex
113 chromosomes (Lenormand, et al. 2020; Lenormand and Roze 2022), we predicted ACRs may be evolving
114 more rapidly than neutrally evolving synonymous sites within coding regions. To survey ACRs, we
115 utilized Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) from liver tissue (Naftaly
116 et al. 2021) and testis tissue. In the liver, we found a total of 1,279 X-linked ACRs. We were able to align
117 948 (74.1%) of these ACRs to orthologous regions on the Y chromosome and the ancestral ninespine
118 stickleback autosome 19. In testis tissue, we identified 896 X-linked ACRs. Compared to the liver, we
119 found far fewer orthologous regions on the Y chromosome and autosome 19 in the ninespine stickleback
120 (118 total aligned; 13.2%), suggesting regulatory regions within testis ACRs may be more rapidly
121 evolving. We estimated divergence between the X and Y chromosomes across ACRs in each tissue.
122 Matching these predictions, among the ACRs we were able to align, testis ACRs had higher divergence
123 between the X and Y chromosomes relative to both synonymous coding region divergence and
124 divergence within liver ACRs (Figure 1). This pattern was consistent across the three evolutionary strata.
125 Although liver divergence within ACRs was not as high as testis ACRs, it was still significantly higher
126 than synonymous coding region divergence in the two oldest strata (one and two) (Figure 1), indicating
127 liver ACRs are also evolving rapidly. In the youngest stratum (three), the liver ACRs were not
128 significantly different than synonymous substitutions. Our results highlight that ACRs are likely under
129 positive selection and this has the strongest effect in ACRs that harbor regulatory elements that are
130 functional in testes.

131

132 **Mutations within accessible regions are mostly accumulating on the Y chromosome**

133 We used the orthologous sequence from the ninespine stickleback fish to infer whether
134 substitutions were accumulating at a higher rate on the X or Y chromosome. ACRs functional in the liver
135 exhibited higher substitution rates on the Y chromosome than on the X chromosome. Y-linked ACRs had
136 significantly higher substitution rates compared to random intergenic regions for the oldest two
137 evolutionary strata (Figure 2, $P < 0.001$; stratum one and two, 10,000 permutations). In contrast to the Y-
138 linked ACRs, we found the X-linked liver ACR sequence divergence was significantly lower than
139 intergenic permutations, consistent with purifying selection operating across all three strata (Figure 2;
140 Stratum one: $P = 0.016$; Stratum two and three: $P < 0.001$; 10,000 permutations). This suggests that ACRs
141 may be functionally constrained more on the X chromosome, compared to Y-linked ACRs. We observed

142 a similar pattern on the Y chromosome for ACRs function in testis tissue. Y-linked ACRs had
143 significantly higher substitution rates in all three evolutionary strata, compared to random intergenic
144 regions (Figure 2; $P < 0.001$, 10,000 permutations). However, we observed X-linked ACRs from testis
145 tissue, also exhibited elevated nucleotide substitutions in the oldest two evolutionary strata (Figure 2; $P <$
146 0.001, 10,000 permutations), while the youngest stratum was significantly lower than intergenic regions
147 (Figure 2; $P < 0.001$, 10,000 permutations). This indicates some ACRs may also be under positive
148 selection on the X chromosome and this occurs in a tissue-specific fashion.

149 Autosomal ACRs have been shown to be under purifying selection in other species (Horvath, et
150 al. 2021). We therefore tested whether threespine stickleback autosomes ACRs also exhibited purifying
151 selection, similar to the pattern we observed on the X chromosome in liver. Consistent with patterns of
152 purifying selection, we found autosomal ACRs exhibited lower sequence divergence compared to random
153 intergenic regions in both liver (Figure 2; $P = 0.002$; 10,000 permutations) and testis tissue (Figure 2; $P <$
154 0.001; 10,000 permutations).

155

156 **Accessible chromatin regions have low nucleotide diversity**

157 We searched for signatures of selection in ACRs from 12 sequenced males from a freshwater
158 population of stickleback fish. Low nucleotide diversity within ACRs relative to neutrally evolving
159 intergenic regions would be indicative of either purifying or positive selection. On autosomes, ACRs have
160 been shown to exhibit lower nucleotide diversity than intergenic regions (Horvath, et al. 2021),
161 suggesting conserved function. We first tested whether we observed a similar pattern among testis and
162 liver tissue ACRs on autosome 18. We found that liver ACRs on chromosome 18 had lower nucleotide
163 diversity compared to intergenic regions (Figure 3; $P < 0.001$, 10,000 permutations), similar to what has
164 been observed in other species. However, we did not detect a significant difference for autosomal testis
165 ACRs (Figure 3; $P = 0.251$, 10,000 permutations). We also did not detect differences in nucleotide
166 diversity between ACRs and intergenic regions on the X or Y chromosomes in liver and testis tissue
167 (Figure 3; all comparisons $P > 0.05$, 10,000 permutations). Combined with low divergence between
168 species, the low nucleotide diversity within ACRs identified from liver tissue indicates some ACRs are
169 under purifying selection, likely due to functional constraints. However, this effect is not consistent across
170 chromosomes or tissues.

171 Among neutrally evolving intergenic sites, we detected major differences among the autosomes
172 and sex chromosomes. Due to a lower effective population size, the Y chromosome is expected to be one-
173 quarter of the nucleotide diversity observed on autosomes, given an equal sex ratio in the population. We

174 found Y chromosome nucleotide diversity within intergenic regions was much lower than expected,
175 compared to autosomes (Y chromosome pi: 0.00013; autosome pi: 0.0050; Figure 3). This suggests that
176 selection acting on linked deleterious or beneficial mutations has lowered nucleotide diversity beyond
177 neutral expectations.

178

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

180 Selection should favor the loss of expression from the Y chromosome in order to silence coding
181 regions that have accumulated deleterious mutations (Orr and Kim 1998; Bachtrog 2006; Lenormand, et
182 al. 2020). To test if Y-linked regulatory divergence was associated with changes in expression, we
183 compared ACR sequence divergence to changes in allelic expression on the X and Y from liver and testis
184 RNA-seq transcriptomes (Peichel, et al. 2020). To maximize the total number of genes for this analysis,
185 we pooled genes across all three evolutionary strata. We found that sequence divergence in liver ACRs,
186 that were proximal to genes, were a predictor of allele-specific expression. We observed higher
187 expression of the X-linked gametolog, relative to the Y-linked gametolog, when ACRs on the Y
188 chromosome had more substitutions (Figure 4A; N = 51*;* liver expression, P = 0.004, Spearman's rank
189 correlation). There was a weaker correlation between expression and substitutions within ACRs on the X
190 chromosome (Figure 4C: N = 51, P = 0.033, Spearman's rank correlation). Combined, these results
191 suggest substitutions on the Y chromosome have a greater effect on increasing X-biased gametolog
192 expression. Although we do not have gene expression from the ninespine stickleback to determine which
193 sex chromosome is deviating from the ancestral expression level, the Y-linked substitutions indicate it is
194 likely due to downregulating the gametolog on the Y chromosome. A similar trend was observed in testis
195 tissue, but the correlations were not significant, likely due to the small number of genes available for the
196 analysis (Figure 4B N = 31, P = 0.111, Spearman's rank correlation).

197

198 **Regulatory evolution is not correlated with coding sequence evolution**

199 The downregulation of Y-linked gametologs could be adaptive if loss of expression follows the
200 accumulation of deleterious coding substitutions (Orr and Kim 1998). We searched for evidence of
201 adaptive silencing on the Y chromosome by comparing the ACR divergence to the ratio of non-
202 synonymous to synonymous substitutions (d_N/d_S) within the coding region of each gene. An adaptive
203 silencing model would be supported if Y-linked gametologs with elevated non-synonymous substitutions
204 relative to synonymous substitutions (d_N/d_S) have elevated regulatory divergence. Among genes shared

205 between the X and Y chromosomes, we found no correlation between ACR nucleotide divergence and
206 d_N/d_S (Supplemental Figure 2; each stratum on X and Y: P > 0.05; Spearman's rank correlation),
207 indicating coding regions with elevated ACR divergence are not more likely to have elevated
208 substitutions in the amino acid sequence. We also examined whether ACR divergence was higher within
209 genes that contained frameshift mutations, nonsense mutations, or in-frame deletions. These types of
210 mutations are more likely to produce a non-functional or sub-optimal peptide. We found no significant
211 difference in the number of ACR substitutions from liver and testis tissue for coding regions that
212 contained frameshift, nonsense mutations, or deletions, compared to coding regions that did not have
213 these putatively deleterious mutations (functional) (Supplemental Figure 3) (all pairwise comparisons P >
214 0.05; Mann-Whitney U test). Overall, we found no evidence of regulatory divergence associated with
215 deleterious coding sequence divergence, suggesting *cis*-regulatory evolution evolves independently of Y-
216 linked coding variation**.**

217

218 **Discussion**

219 Gene regulation evolves rapidly on degenerating Y chromosomes. Our understanding of this
220 process has largely been informed by characterizing gametolog-specific gene expression from
221 degenerating sex chromosomes (Muyle, et al. 2012; White, et al. 2015; Beaudry, et al. 2017; Muyle, et al.
222 2018; Martin, et al. 2019; Veltsos, et al. 2019). Over time, gametolog expression is generally lost across
223 most genes on the Y chromosome. With the growing number of Y chromosome reference assembles and
224 sequencing methods to quickly profile functional regulatory regions across the genome, it has become
225 feasible to explore the molecular evolution of these regions that lead to altered expression patterns. Here,
226 we observed elevated nucleotide substitutions within regulatory elements across the threespine
227 stickleback Y chromosome, consistent with positive selection driving this process. We found that
228 increased rates of *cis*-regulatory divergence were associated with X-biased expression of XY gametologs,
229 suggesting loss of expression from the Y chromosome. Without liver and testis gene expression from the
230 ninespine stickleback as an ancestral comparison, we are unable to confirm that X-biased gametolog
231 expression is a result of downregulation of the Y chromosome or upregulation of the X chromosome.
232 However, previous sequencing of a brain transcriptome showed expression patterns consistent with loss
233 of Y expression rather than gain of X expression (White, et al. 2015). It is therefore likely that the X-
234 biased expression we observed in liver and testis tissue is also through loss of Y chromosome expression,
235 similar to what has been observed in other species with degenerating Y chromosomes (Muyle, et al. 2012;
236 Beaudry, et al. 2017; Wei and Bachtrog 2019).

237 Simulations indicate rapid expression loss from the Y chromosome is driven by the divergence of
238 *cis*-regulatory elements on both the X and Y chromosomes (Lenormand, et al. 2020; Lenormand and
239 Roze 2022). Mutations initially accumulate in regulatory elements within the non-recombining region of
240 the Y chromosome, which can lead to some expression loss. As additional deleterious mutations arise in
241 the coding sequence of genes, positive selection will favor the accumulation of additional regulatory
242 substitutions to silence expression of the Y-linked gametolog. Non-synonymous mutations are
243 accumulating at a higher rate across the threespine stickleback Y chromosome (White, et al. 2015;
244 Peichel, et al. 2020). However, we found that the number of nucleotide substitutions within regulatory
245 regions was not correlated with deleterious mutations within neighboring coding regions. Our results
246 suggest that adaptive regulatory evolution may happen independently of mutations within the amino acid
247 sequence. Our results complement previous findings that have compared coding sequence evolution to
248 gametolog expression. The level of expression from the Y chromosome is often not correlated with the
249 overall number of deleterious mutations within coding regions (i.e., gametologs with many deleterious
250 mutations are not necessarily expressed at lower levels from the Y chromosome) (Bachtrog 2006;
251 Bachtrog, et al. 2008; Beaudry, et al. 2017). It is also possible that we did not capture the complete set of
252 regulatory elements important for Y expression in our ATAC-seq datasets. A correlation between
253 nucleotide substitutions in regulatory regions and deleterious mutations in coding regions may be
254 observed if ACRs and gene expression are profiled from additional tissues. In addition, we assigned
255 ACRs to coding sequences based on overall proximity to a gene. Although this method of annotating
256 ACRs is commonly used to assign regulatory regions to genes (Connelly, et al. 2014; Alexandre, et al.
257 2018; Ricci, et al. 2019), it is possible that some ACRs interact with other genes through long range
258 interactions based on chromatin configuration (Mifsud, et al. 2015; Schoenfelder and Fraser 2019),
259 reducing our ability to detect an association between deleterious mutations in coding regions and
260 nucleotide substitutions in ACRs. Finally, the number of substitutions within an ACR may not be a fully
261 accurate predictor of the overall ability to silence a given gametolog on the Y chromosome. If only a
262 small number of substitutions are needed to ablate transcription factor binding within an ACR, we would
263 not expect to see a strong correlation between the number of deleterious mutations within coding
264 sequences and the number of substitutions within ACRs. Additional functional work will be necessary to
265 explore these alternatives.

266 Degeneration of *cis*-regulatory elements on the Y chromosome and loss of expression should
267 select for upregulation of the gametolog on the X chromosome to compensate for dosage loss
268 (Lenormand, et al. 2020). In this scenario, nucleotide substitutions should also accumulate within ACRs
269 on the X chromosome. In contrast to this hypothesis, we found limited evidence for widespread positive
270 selection across X-linked regulatory regions. We found elevated rates of X-linked ACR divergence in

271 testis, but this did not extend to ACRs from the liver. In many cases, ACRs on the X chromosome
272 actually exhibited signatures of purifying selection, with substitution rates lower than random intergenic
273 regions. This pattern was also observed on the autosomes. Chromosome-wide dosage compensation does
274 not occur in the threespine stickleback fish. Genes that have completely lost the Y-linked gametolog
275 exhibit half expression in males, relative to females (White, et al. 2015). The lack of positive selection
276 among X-linked ACRs is consistent with most genes on the X and Y chromosomes not being strongly
277 dosage sensitive. In this case, most gametologs on the Y chromosome could be lost without consequence.
278 Even in the absence of strong stabilizing selection to maintain dosage balance, *cis*-regulatory
279 degeneration is still predicted to occur rapidly (Lenormand, et al. 2020). The Y-specific positive selection
280 in ACRs we detected support this scenario.

281 Low nucleotide diversity within a population is a signature of positive or purifying selection
282 (Nielsen 2005). We explored whether there was an effect of positive selection in ACRs on nucleotide
283 diversity within a freshwater population of threespine stickleback fish. Although the elevated nucleotide
284 divergence between the X and Y we observed in ACRs is a signature of positive selection, we did not
285 detect corresponding signatures of reduced nucleotide diversity within Y-linked ACRs. However, Y
286 chromosomes often have much lower diversity overall, relative to neutral expectations (Lawson Handley,
287 et al. 2006; Wilson Sayres, et al. 2014). In such a case, it is difficult to identify the target of any selective
288 sweeps. Similar to humans (Wilson Sayres, et al. 2014), our findings revealed that nucleotide diversity on
289 the threespine stickleback Y chromosome is much lower than neutral expectations. Low Y-linked
290 diversity has been attributed to purifying selection removing deleterious mutations and affecting linked
291 neutral sites throughout the entire non-recombining region (Wilson Sayres, et al. 2014; Wilson Sayres
292 2018). We found ACRs exhibited high divergence between species, but exhibited low nucleotide diversity
293 within populations, similar to intergenic regions throughout the chromosome. These results suggest
294 positive selection within ACRs could also be an important driver of reducing nucleotide diversity on the
295 Y chromosome within populations. One important consideration is whether sex-biased demography may
296 be affecting Y-linked diversity. If males are more variable in reproductive success than females, this
297 could lower the expected effective population size of the Y chromosome, reducing nucleotide diversity.
298 Little is known about the operating sex-ratios of stickleback fish. Some populations exhibit sex ratios that
299 are female biased (Rollins, et al. 2017). This could contribute to the low nucleotide diversity we observed.
300 However, simulations in human populations revealed that even drastic shifts in sex ratio could not entirely
301 explain the low diversity observed on the Y chromosome (Wilson Sayres, et al. 2014).

302 Although expression is lost from most genes across the Y chromosome, the coding sequence of
303 some genes are maintained through purifying selection. These genes tend to be enriched for

304 haploinsufficiency phenotypes and presumably cannot be lost from the Y chromosome (Bellott, et al.
305 2014; Bellott, et al. 2017; Peichel, et al. 2020; Bellott and Page 2021). For this category of genes, if *cis*-
306 regulatory elements are accumulating deleterious mutations, ancestral expression patterns could be
307 maintained by gaining new binding sites that maintain expression. This would also result in signatures of
308 increased substitution rates relative to other regions. This type of compensatory evolution has been
309 proposed for functionally critical genes in mammals (Chaix, et al. 2008; Vermunt, et al. 2016), and
310 *Drosophila* (Arnold, et al. 2014). Functional analysis of what transcription binding sites are affected by
311 Y-linked mutations could help identify which mutations are leading to down-regulation of gametologs
312 compared to those that restore ancestral expression.

313 **Conclusion**

314 Together, our results provide evidence of positive selection driving accelerated rates of nucleotide
315 substitution in *cis*-regulatory elements. Consistent signatures across all three evolutionary strata indicate
316 that *cis*-regulatory evolution can proceed rapidly following the suppression of recombination, leading to
317 reduced gene expression from the Y chromosome. Improvements in functional annotations of regulatory
318 regions as well as an ever-growing collection of high-quality Y and W assemblies will allow continued
319 empirical testing of new regulatory models of sex chromosome evolution.

320

321 **Materials and Methods**

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

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

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

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

358 We explored overall concordance between each replicate by assaying whether ATAC-seq reads
359 were enriched around transcription start sites in the two testis and two liver samples. Iso-Seq long read
360 sequencing was previously conducted in threespine stickleback fish to curate accurate transcription start
361 site annotations across multiple tissues (Naftaly, et al. 2021). We used deepTools (v3.5.1) computeMatrix
362 reference-point (Ramírez, et al. 2016) on the ATAC-seq alignments to assay read depth 3kb around the
363 complete set of annotated transcription start sites. We plotted enrichment of ATAC-seq reads around TSS
364 and compared to RNA-seq expression using deepTools plotHeatmap with default settings, except that we
365 used the -sortUsingsamples to sort the regions by expression. We found replicates across both tissues
366 were highly concordant in which transcription start sites were accessible (Supplemental Figure 1). We
367 also found accessible transcription start sites were a good predictor of gene expression in each tissue

368 (Supplemental Figure 1). Because both replicates in each tissue were highly concordant, we pooled the
369 aligned reads for each tissue to maximize read depth for all the analyses. ACRs were called for each
370 tissue sample using MACS2 (v2.2.7.1) callpeak with the –keep--dup all parameter, and read depth was
371 normalized with Tn5 control sequencing with the -c parameter (Zhang, et al. 2008).

372 We mapped the nucleotide sequence of each ACR on the X chromosome to homologous regions
373 on the Y chromosome using previously generated alignment coordinates (Peichel, et al. 2020). We
374 identified orthologous regions between the X chromosome, Y chromosome, and orthologous autosome
375 (chromosome 19) from the ninespine stickleback fish *(Pungitius pungitius*; (Varadharajan, et al. 2019),
376 2019) using BLAST+ (blastn v. 2.11.0) with default blastn parameters, with -perc_identity set to 75.
377 (Camacho, et al. 2009). To ensure high quality alignments, we filtered for uniquely mapping alignments
378 that also had a bit score greater than 100. We used BLAST+ alignments to extract sequence to create
379 multiple sequence alignments for downstream analysis. We aligned the threespine stickleback X and Y
380 ACR sequences to the ninespine stickleback autosome sequence using MUSCLE (v 3.8.1551) (Edgar
381 2004) with default parameters. We used MUSCLE to create a multiple sequence alignment between the X
382 sequence, Y sequence, and the sequence from the orthologous autosome in order to identify X- and Y-
383 specific variants. We called single nucleotide variants using a custom python script modified from
384 msa2snps.py (Junli Zhang, https://github.com/pinbo/msa2snp). We calculated the reported divergence
385 rate by dividing the number of variants by the number of the aligned sites.

386 We compared ACR divergence on the sex chromosomes to ACR divergence on autosome 18, a
387 chromosome similar in length to the X chromosome. We identified orthologous regions between the
388 threespine stickleback autosome 18 and autosome 18 from the ninespine stickleback fish using BLAST+
389 (blastn v. 2.11.0) (Camacho, et al. 2009) as previously described. We called variants between autosomal
390 regions in a similar manner as the sex chromosomes, by creating a pairwise alignment using MUSCLE (v
391 3.8.1551) with default parameters and called single nucleotide variants using the same python script as
392 above. We calculated a divergence rate by dividing the number of single nucleotide variants by the size of
393 the aligned region.

394 Divergence within ACRs was compared with the divergence of coding regions of neighboring
395 genes. We identified the closest gene to each ACR by running annotatepeaks.pl in homer (v 4.11) (Heinz,
396 et al. 2010) using default settings. Per default settings, ACRs were assigned to the closest gene within
397 50kb of the TSS. For coding divergence comparisons, we used previously reported estimates of
398 synonymous (d_S) and non-synonymous (d_N) divergence of gametologs between the X and Y
399 chromosomes (Peichel et al., 2020). Coding regions with d_N/d_S ratios of 99 were omitted.

400

401 **Estimating a neutral substitution rate**

402 We used randomly drawn intergenic regions throughout the X chromosome, Y chromosome, or
403 chromosome 18 to estimate a neutral substitution rate. Regulatory regions tend to be GC-rich, which can
404 be prone to higher mutation rates through mechanisms like GC-biased gene conversion or spontaneous
405 deamination of methylated cytosines (Nesta, et al. 2021). We therefore GC-matched the randomly drawn
406 intergenic regions with the GC content of ACRs. We calculated the GC percentage of ACRs using a
407 modified Perl script (countbp.pl, Nicholas Navin,
408 http://www.navinlab.com/bioperl/bioperl/gc_content.html). For each ACR, a random intergenic region
409 was drawn from the X chromosome equal in length and with a GC content within 2.0%. Intergenic
410 regions were defined as any region that that falls within annotated functional regions from a combination
411 of Ensembl annotations (Cunningham, et al. 2021), Isoseq transcripts (Naftaly, et al. 2021), Y
412 chromosome annotations (Peichel, et al. 2020), repetitive elements (Peichel, et al. 2020; Nath, et al. 2021)
413 and ACRs from this study. This was performed for each stratum individually using bedtools shuffle –I
414 ACRs.bed -incl XYalignment.bed -excl annotations.bed. We extracted 10,000 sets of GC-matched
415 intergenic regions from each of the three evolutionary strata on the threespine stickleback sex
416 chromosomes (set size same as number of Liver ACRs: stratum one: 442, stratum two: 228, stratum three:
417 199; Testis ACRs: stratum one: 30, stratum two: 37, stratum three: 51). Due to a limited number of
418 intergenic sites shared between the sex chromosomes, we sampled with replacement to align a sufficient
419 number of intergenic regions. We identified the orthologous regions to the ninespine genome assembly
420 using BLAST+ (blastn v2.11.0) (Camacho, et al. 2009). We filtered for alignments that mapped uniquely
421 to chromosome 19 in the ninespine stickleback fish, had an alignment length equal to 75% of the
422 threespine query sequence, and had a bit-score greater than 100. We generated multiple sequence
423 alignments using MUSCLE (v 3.8.1551), for each intergenic permutation and calculated X and Y specific
424 substitution rate as previously described.

425

426 **Estimating within population nucleotide diversity**

427 We used whole-genome short-read sequencing of 12 males from the Lake Washington population
428 (Washington, USA; NCBI SRA SRP137809) to estimate nucleotide diversity across the sex
429 chromosomes. The raw reads were trimmed with Trimmomatic (v. 0.39) (Bolger, et al. 2014), using a
430 sliding window of four bases, trimming the remainder of the read when the average quality within a
431 window dropped below 15. The leading and trailing base pairs below quality three of every read were

432 removed along with any residual adapter sequence. After trimming, any reads below a minimum length of
433 36 were discarded. We aligned the trimmed reads using Bowtie2 (v. 2.4.5) with default parameters
434 (Langmead and Salzberg 2012). We also marked PCR duplicates by using the MarkDuplicates function in
435 Picard (v. 2.26.10) (https://github.com/broadinstitute/picard). SNP genotyping was conducted following
436 the GATK (v. 4.2.5.0) best practices for germline short variant discovery. Variants were called per sample
437 using HaplotypeCaller in GVCF mode. We then joint-called variants using GenotypeGVCFs using a
438 genomics database created by GenomicsDBImport.

439 We only considered biallelic SNPs in our estimates of nucleotide diversity and we filtered for
440 high quality genotypes using several different methods. On the sex chromosomes, sites should only have
441 hemizygous genotypes in males in the non-crossover region outside of the pseudoautosomal region. Any
442 sites that are heterozygous would be caused by errors in read alignment. We therefore did not consider
443 sites that were heterozygous in the non-crossover region on the X or Y chromosomes. We also filtered out
444 sites that exhibited too low or high of read depth, which would be indicative of alignment errors. To do
445 this, we did not consider sites that were less than one-half or greater than double the median read depth of
446 each chromosome (X and Y chromosomes: positions were retained if the read depth was between 3.5 and
447 14; chromosome 18: positions were retained if the read depth was between 6.5 and 26). Read filtering was
448 conducted using custom Perl scripts.

449 Nucleotide diversity (pi) was estimated on a per-site basis (--site-pi) across each chromosome
450 using vcftools (v. 0.1.16) (Danecek, et al. 2011)). Nucleotide diversity for ACRs was computed as the
451 average diversity across all sites within ACRs, including non-variant (zero pi) sites that were not removed
452 by read depth filters. We compared nucleotide diversity within ACRs with nucleotide diversity in
453 randomly drawn intergenic sites. Intergenic sites were defined as previously described. Intergenic sites
454 were drawn with replacement equal to the total number of sites within ACRs on each chromosome
455 (10,000 total permutations).

456

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

458 We measured allele-specific expression of transcripts on the X and Y chromosome using three
459 replicate liver samples (NCBI BioProject PRJNA591630) and three replicate testis samples (NCBI
460 BioProject PRJNA591630). We trimmed the RNA-seq reads with Trimmomatic (v. 0.36) (Bolger, et al.
461 2014) using a sliding window of 4 bases, trimming the remainder of the read when the average quality
462 within a window dropped below 20 (SLIDINGWINDOW:4:20). Residual sequencing adapters were also
463 removed using ILLUMINACLIP. We aligned RNA-seq reads using Tophat2 (v2.2.1) with default

464 parameters (Kim, et al. 2013). We filtered for alignments with a map quality greater than 25 using
465 SAMtools (v1.14) (Li, et al. 2009) to identify reads that map uniquely on the X and Y allele. Similar map
466 quality filters have been used to distinguish short reads between sex chromosomes in *Drosophila* and
467 stickleback fish (Ellison and Bachtrog 2019; Peichel, et al. 2020) and between sexes with heteromorphic
468 sex chromosomes in guppies (Kirkpatrick, et al. 2021). Read counts were obtained using htseq-count (v
469 0.9.1) (Anders, et al. 2015). Default parameters were used with the addition of --stranded=no and --
470 nonunique all to maximize our ability to count reads on the Y chromosome. We used a custom GTF file
471 for htseq-count that included previously determined start and end sites of all Ensembl predicted
472 transcripts that are shared between the X and Y chromosomes (Peichel, et al. 2020). Reads were counted
473 on the X and Y across the entire transcript as a feature. Transcripts were removed from the analysis if
474 they had an RNA expression count of zero in all samples. For each tissue, we calculated the average read
475 count for each gametolog across the three tissue replicates. We calculated the allele-specific expression as
476 $\log_2$(X transcript average / Y transcript average).

477

478 **Data availability**

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

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

483 **Acknowledgements**

484 This work was supported by a National Science Foundation MCB 1943283 to M.A.W.

485 **Declaration of interests**

486 The authors declare no conflicts of interests.

487

488 **Figure legends**

489

490 **Figure 1. XY divergence within accessible chromatin regions and coding regions.** ACRs show
491 elevated divergence between the X and Y, relative to nonsynonymous and synonymous substitutions in
492 coding regions. Protein-coding divergence was estimated between the X and Y chromosomes at all shared
493 gametologs. Accessible Chromatin Regions (ACR) were identified on the X chromosome and mapped to
494 the Y chromosome. Substitutions within ACRs were summed and divided by the total length of the ACR
495 to calculate divergence. Stratum one: 12 gametologs, 442 liver ACRs, 30 testis ACRs; Stratum two: 71
496 gametologs, 228 liver ACRs, 37 testis ACRs; Stratum three: 46 gametologs, 199 liver ACRs 51 testis
497 ACRs. Divergence within each category was compared using a pairwise Mann-Whitney U test, adjusting
498 for multiple comparisons. The letters above each plot indicate the significant differences between sites
499 within each stratum ($P < 0.01$).

500

501 **Figure 2. Accessible chromatin divergence between threespine and ninespine stickleback.** Y-linked
502 ACR divergence is elevated over neutrally evolving intergenic regions. X and Y ACR divergence was
503 compared to randomly drawn intergenic regions matched for GC content (10,000 permutations). The
504 average ACR divergence of each stratum is shown by the dotted line (stratum 1: pink, stratum 2: green,
505 and stratum 3: blue). ACRs were identified on the X chromosome and mapped to the Y chromosome and
506 ninespine stickleback *(Pungititus pungititus*) autosome (chromosome 19) in liver and testis. Autosome
507 divergence is for chromosome 18. X- and Y- specific within ACRs were summed and divided by the total
508 length of the ACR to calculate divergence. Stratum one: 442 liver ACRs and 30 testis ACRs; stratum two:
509 228 liver ACRs and 37 testis ACRs; stratum three: 129 liver ACRs and 46 testis ACRs; autosomal: 1056
510 liver ACRs and 305 testis ACRs.

511

512 **Figure 3. Regulatory nucleotide diversity (pi) compared to intergenic regions.** Nucleotide diversity
513 within ACRs on the X and Y chromosome are not different than randomly drawn intergenic regions
514 (10,000 permutations). ACRs were identified on the X chromosome and mapped to the Y chromosome.
515 Nucleotide diversity of ACRs is also shown on chromosome 18. Average ACR nucleotide diversity for
516 each chromosome is shown by the dotted lines. Y chromosome ACR nucleotide diversity is not different
517 from intergenic permutations in liver ($P = 0.623$) or testis ($P = 0.374$). X chromosome ACR nucleotide
518 diversity is not different than intergenic regions in liver ($P = 0.1985$) or testis ($P = 0.147$). Autosomal
519 ACRs exhibited significantly lower diversity compared to intergenic permutations from liver, not from
520 testis (Liver: $P < 0.001$, Testis $P = 0.252$). Stratum one: 442 liver ACRs and 30 testis ACRs; stratum two:
521 228 liver ACRs and 37 testis ACRs; stratum three: 129 liver ACRs and 46 testis ACRs; autosomal: 1056
522 liver ACRs and 305 testis ACRs.

523

524 **Figure 4**. **Gametolog expression level compared to ACR divergence.** ACR divergence on the X and Y
525 chromosomes is correlated with X-biased gene expression in both liver and testis tissues. RNA-seq
526 transcript counts uniquely mapped to the X and Y chromosome were quantified to determine allele-
527 specific expression for all gametologs. Gametolog expression was compared to average divergence of Y-
528 specific (A and B) and X-specific (C and D) mutations for ACRs within 50 kb of expressed genes. (A). Y
529 liver: $R= 0.40$, $P = 0.004$, $N = 65$; (B.) Y Testis: $R= 0.34$, $P = 0.008$, $N = 65$; (C). X Liver: $R = 0.30$, $P =$
530 0.033, $N = 77$; (D). X Testis: $R = 0.25$, 0.061, $N = 77$.

531

532 **Supplemental Figure 1. High concordance of chromatin accessibility around transcription start site**
533 **of expressed genes**. Transcription start sites (TSSs) for all genes were ordered by RNA-seq read depth.
534 ATAC-seq read depth was quantified ±3 kb of TSSs. Similar ATAC-seq read depth was observed
535 between replicates. ATAC-seq depth across replicates is a strong predictor of expression from RNA-seq
536 from same tissue type.

537

538 **Supplemental Figure 2. d_N / d_S ratios of coding regions compared to ACR divergence.** There is no
539 correlation between d_N / d_S and ACR divergence using a Spearman's Rank Correlation using ACRs from
540 liver (A and B) or testis (C and D) tissue. Liver: (**A**) Y divergence: (Stratum one (Blue): R=-0.12 P =
541 0.230, N = 30, Stratum two (Yellow): R = 0.2, P = 0.430, N=61; Stratum three (Grey): R = 0.11, P =
542 0.770 N=126. (**B**) X divergence: (Stratum one (Blue): R = -0.080, P = 0.420, N=30; Stratum two
543 (Yellow): R = 0.032, P = 0.9, N = 61; Stratum three (Grey): R = -0.054, P = 0.890, N = 126. Similar
544 results in testis: (**C**) Y divergence: (Stratum one (Blue): R = -0.020, P = 0.650, N=14; Stratum two
545 (Yellow): R = -0.430, P = 0.160, N = 12; Stratum three (Grey): R = -0.140, P = 0.480, N = 17. (**D**) X
546 divergence: (Stratum one (Blue): R = 0.520, P = 0.120, N=14; Stratum two (Yellow): R = -0.360, P =
547 0.210, N = 12; Stratum three (Grey): R = -0.020, P = 0.670, N = 17.

548

549 **Supplemental Figure 3. ACR divergence compared to functional states of Y-linked gametologs.**
550 ACR divergence was not significantly different among gametologs that contained mutations that
551 potentially impede function (i.e., premature stop codons, frameshifts, deleted nucleotides) or are
552 functional. Y-specific functional mutations were identified by comparing coding sequence between the Y
553 chromosome and the outgroup autosomal coding sequence in the ninespine stickleback. ACRs were
554 assigned to the closest gametolog within 50kb. P > 0.05 for all pairwise comparisons for both liver (A and
555 C) and testis (B and D) for regulatory divergence on the X and Y.

556

557

558 **References**

559

560 Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, Sullivan AM, Bemm F,
561 Jolic D, Arsovski AA, et al. 2018. Complex Relationships between Chromatin Accessibility, Sequence
562 Divergence, and Gene Expression in Arabidopsis thaliana. Molecular Biology and Evolution 35:837-854.
563 Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing
564 data. Bioinformatics 31:166-169.
565 Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. 2014. Quantitative
566 genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation
567 and turnover during cis-regulatory evolution. Nature Genetics 46:685-692.
568 Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. 2013. RNA
569 sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and
570 allows comprehensive annotation of the W-chromosome. Genome Biology 14:R26.
571 Bachtrog D. 2006. Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. Current
572 Biology 16:1694-1699.
573 Bachtrog D. 2008. The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration.
574 Genetics 179:1513-1525.
575 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome
576 degeneration. Nature Reviews Genetics 14:113-124.
577 Bachtrog D, Hom E, Wong KM, Maside X, De Jong P. 2008. Genomic degradation of a young Y
578 chromosome in Drosophila miranda. Genome Biology 9:R30.
579 Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I,
580 Ming R, et al. 2014. Sex Determination: Why So Many Ways of Doing It? PLoS Biology 12:e1001899.
581 Beaudry FEG, Barrett SCH, Wright SI. 2017. Genomic Loss and Silencing on the Y Chromosomes of
582 Rumex. Genome Biology and Evolution 9:3345-3355.
583 Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T,
584 Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators.
585 Nature 508:494-499.
586 Bellott DW, Page DC. 2021. Dosage-sensitive functions in embryonic development drove the survival of
587 genes on sex-specific chromosomes in snakes, birds, and mammals. Genome Research 31:198-210.
588 Bellott DW, Skaletsky H, Cho T-J, Brown L, Locke D, Chen N, Galkina S, Pyntikova T, Koutseva N, Graves T,
589 et al. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators.
590 Nature Genetics 49:387-394.
591 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
592 Bioinformatics 30:2114-2120.
593 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for
594 fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome
595 position. Nature Methods 10:1213-1218.
596 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
597 architecture and applications. BMC Bioinformatics 10:421.
598 Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA. 2008. Evolution of Primate Gene Expression: Drift
599 and Corrective Sweeps? Genetics 180:1379-1389.
600 Chang C-H, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. 2022. Unique structure and positive
601 selection promote the rapid divergence of Drosophila Y chromosomes. eLife 11.
602 Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. Proceedings of
603 the National Academy of Sciences 75:5618-5622.

604 Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philosophical Transactions
605 of the Royal Society of London. Series B: Biological Sciences 355:1563-1572.
606 Connelly CF, Wakefield J, Akey JM. 2014. Evolution and Genetic Architecture of Chromatin Accessibility
607 and Function in Yeast. PLoS Genetics 10:e1004427.
608 Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, Austine-Orimoloye O,
609 Azov Andrey G, Barnes I, Bennett R, et al. 2021. Ensembl 2022. Nucleic Acids Research 50:D988-D995.
610 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT,
611 Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.
612 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
613 Acids Research 32:1792-1797.
614 Ellison C, Bachtrog D. 2019. Recurrent gene co-amplification on Drosophila X and Y chromosomes. PLoS
615 Genetics 15:e1008251.
616 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010.
617 Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements
618 Required for Macrophage and B Cell Identities. Molecular Cell 38:576-589.
619 Horvath R, Josephs EB, Pesquet E, Stinchcombe JR, Wright SI, Scofield D, Slotte T. 2021. Selection on
620 Accessible Chromatin Regions in Capsella grandiflora. Molecular Biology and Evolution 38:5563-5575.
621 Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki
622 C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y
623 chromosomes. Nature 483:82-86.
624 Hughes JF, Skaletsky H, Pyntikova T, Graves TA, Van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke
625 DP, Friedman C, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in
626 structure and gene content. Nature 463:536-539.
627 Hughes JF, Skaletsky H, Pyntikova T, Koutseva N, Raudsepp T, Brown LG, Bellott DW, Cho T-J, Dugan-
628 Rocha S, Khan Z, et al. 2020. Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in
629 mammalian lineages. Genome Research 30:1716-1726.
630 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of
631 transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:R36.
632 Kirkpatrick M, Sardell JM, Pinto BJ, Dixon G, Peichel CL, Schartl M. 2021. Evolution of the canonical sex
633 chromosomes of the guppy and its relatives. G3 Genes|Genomes|Genetics.
634 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
635 Lawson Handley LJ, Berset-BräNdli L, Perrin N. 2006. Disentangling Reasons for Low Y Chromosome
636 Variation in the Greater White-Toothed Shrew (Crocidura russula). Genetics 173:935-942.
637 Lenormand T, Fyon F, Sun E, Roze D. 2020. Sex Chromosome Degeneration by Regulatory Evolution.
638 Current Biology 30:3001-3006.e3005.
639 Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual
640 dimorphism. Science 375:663-666.
641 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The
642 Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
643 Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. 2017. Combining ATAC-seq with nuclei sorting
644 for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Res 45:e41.
645 Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. 2019. The prevalence, evolution and
646 chromatin signatures of plant regulatory elements. Nature Plants 5:1250-1259.
647 Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. Nature
648 Communications 8.
649 Mahajan S, Wei KHC, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young Drosophila Y
650 chromosome using single-molecule sequencing and chromatin conformation capture. PLoS Biology
651 16:e2006348.

652 Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GAB, Touzet P. 2019. Evolution of
653 Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination Systems.
654 Genome Biology and Evolution 11:350-361.
655 Meisel RP, Malone JH, Clark AG. 2012. Disentangling the relationship between sex-biased gene
656 expression and X-linkage. Genome Research 22:1255-1265.
657 Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey
658 W, Ewels PA, et al. 2015. Mapping long-range promoter contacts in human cells with high-resolution
659 capture Hi-C. Nature Genetics 47:598-606.
660 Muller HJ. 1918. Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal
661 Factors. Genetics 3:422-499.
662 Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid De Novo Evolution of X
663 Chromosome Dosage Compensation in Silene latifolia, a Plant with Young Sex Chromosomes. PLoS
664 Biology 10:e1001308.
665 Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Hobza R, Picard F, Widmer A,
666 et al. 2018. Genomic imprinting mediates dosage compensation in a young plant XY system. Nature
667 Plants 4:677-680.
668 Naftaly AS, Pau S, White MA. 2021. Long-read RNA sequencing reveals widespread sex-specific
669 alternative splicing in threespine stickleback fish. Genome Research:gr.274282.274120.
670 Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using
671 long-read sequencing. G3 Genes|Genomes|Genetics 11.
672 Nesta AV, Tafur D, Beck CR. 2021. Hotspots of Human Mutation. Trends in Genetics 37:717-729.
673 Nielsen R. 2005. Molecular Signatures of Natural Selection. Annual Review of Genetics 39:197-218.
674 Orr HA, Kim Y. 1998. An Adaptive Hypothesis for the Evolution of the Y Chromosome. Genetics
675 150:1693-1698.
676 Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, Grimwood J, Schmutz J, Myers RM,
677 Kingsley DM, et al. 2020. Assembly of the threespine stickleback Y chromosome reveals convergent
678 signatures of sex chromosome evolution. Genome Biology 21.
679 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
680 Bioinformatics 26:841-842.
681 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016.
682 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research
683 44:W160-W165.
684 Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-
685 Tatché M, et al. 2019. Widespread long-range cis-regulatory elements in the maize genome. Nature
686 Plants 5:1237-1249.
687 Rodríguez Lorenzo JL, Hobza R, Vyskot B. 2018. DNA methylation and genetic degeneration of the Y
688 chromosome in the dioecious plant Silene latifolia. BMC Genomics 19.
689 Rollins JL, Chiang P, Waite JN, Von Hippel FA, Bell MA. 2017. Jacks and jills: Alternative life-history
690 phenotypes and skewed sex ratio in anadromous threespine stickleback (Gasterosteus aculeatus).
691 Evolutionary Ecology Research 18:363-382.
692 Schoenfelder S, Fraser P. 2019. Long-range enhancer–promoter contacts in gene expression control.
693 Nature Reviews Genetics 20:437-455.
694 Shaw DE, White MA. 2022. The evolution of gene regulation on sex chromosomes. Trends in Genetics.
695 Singh ND, Koerich LB, Carvalho AB, Clark AG. 2014. Positive and Purifying Selection on the Drosophila Y
696 Chromosome. Molecular Biology and Evolution 31:2612-2623.
697 Soh YQ, Alföldi J, Pyntikova T, Laura, Graves T, Patrick, Robert, Kremitzki C, Koutseva N, Jacob, et al.
698 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on
699 Both Sex Chromosomes. Cell 159:800-813.

700 Tomaszkiewicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D, O'Brien PCM,
701 Chikhi R, Ryder OA, et al. 2016. A time- and cost-effective strategy to sequence mammalian Y
702 Chromosomes: an application to the de novo assembly of gorilla Y. Genome Research 26:530-540.
703 Varadharajan S, Rastas P, Löytynoja A, Matschiner M, Calboli FCF, Guo B, Nederbragt AJ, Jakobsen KS,
704 Merilä J. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome.
705 Genome Biology and Evolution.
706 Veltsos P, Ridout KE, Toups MA, González-Martínez SC, Muyle A, Emery O, Rastas P, Hudzieczek V, Hobza
707 R, Vyskot B, et al. 2019. Early Sex-Chromosome Evolution in the Diploid Dioecious Plant Mercurialis
708 annua. Genetics 212:815-835.
709 Vermunt MW, Tan SC, Castelijns B, Geeven G, Reinink P, De Bruijn E, Kondova I, Persengiev S, Bontrop R,
710 Cuppen E, et al. 2016. Epigenomic annotation of gene regulatory alterations during evolution of the
711 primate brain. Nature Neuroscience 19:494-503.
712 Wei KHC, Bachtrog D. 2019. Ancestral male recombination in Drosophila albomicans produced
713 geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. PLoS
714 Genetics 15:e1008502.
715 White MA, Kitano J, Peichel CL. 2015. Purifying Selection Maintains Dosage-Sensitive Genes during
716 Degeneration of the Threespine Stickleback Y Chromosome. Molecular Biology and Evolution 32:1981-
717 1995.
718 Wilson Sayres MA. 2018. Genetic Diversity on the Sex Chromosomes. Genome Biology and Evolution
719 10:1064-1078.
720 Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural Selection Reduced Diversity on Human Y
721 Chromosomes. PLoS Genetics 10:e1004064.
722 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li
723 W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9:R137.

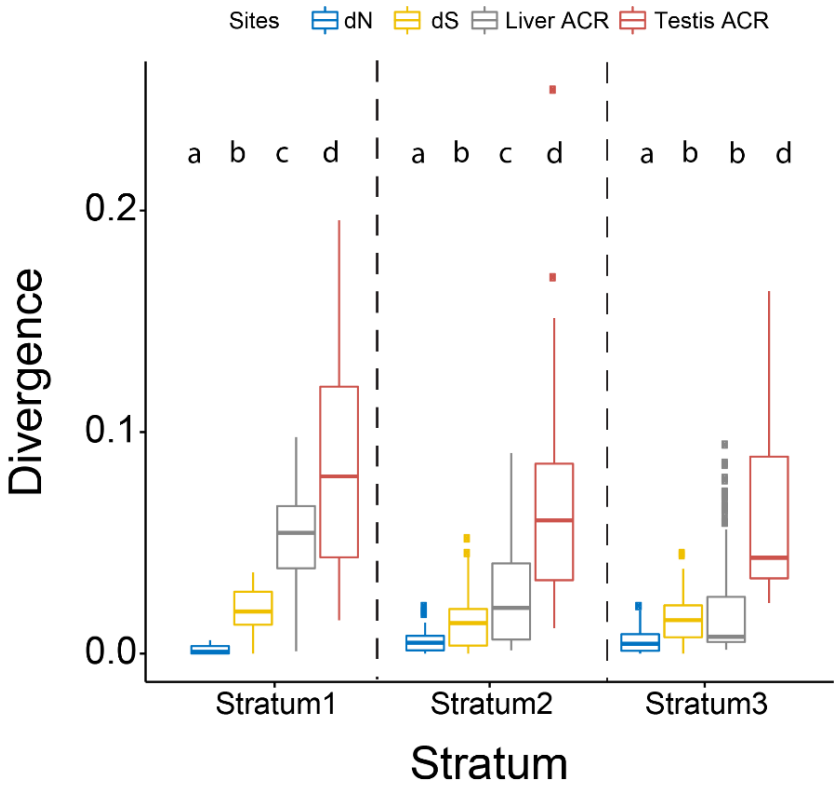
724

725

726

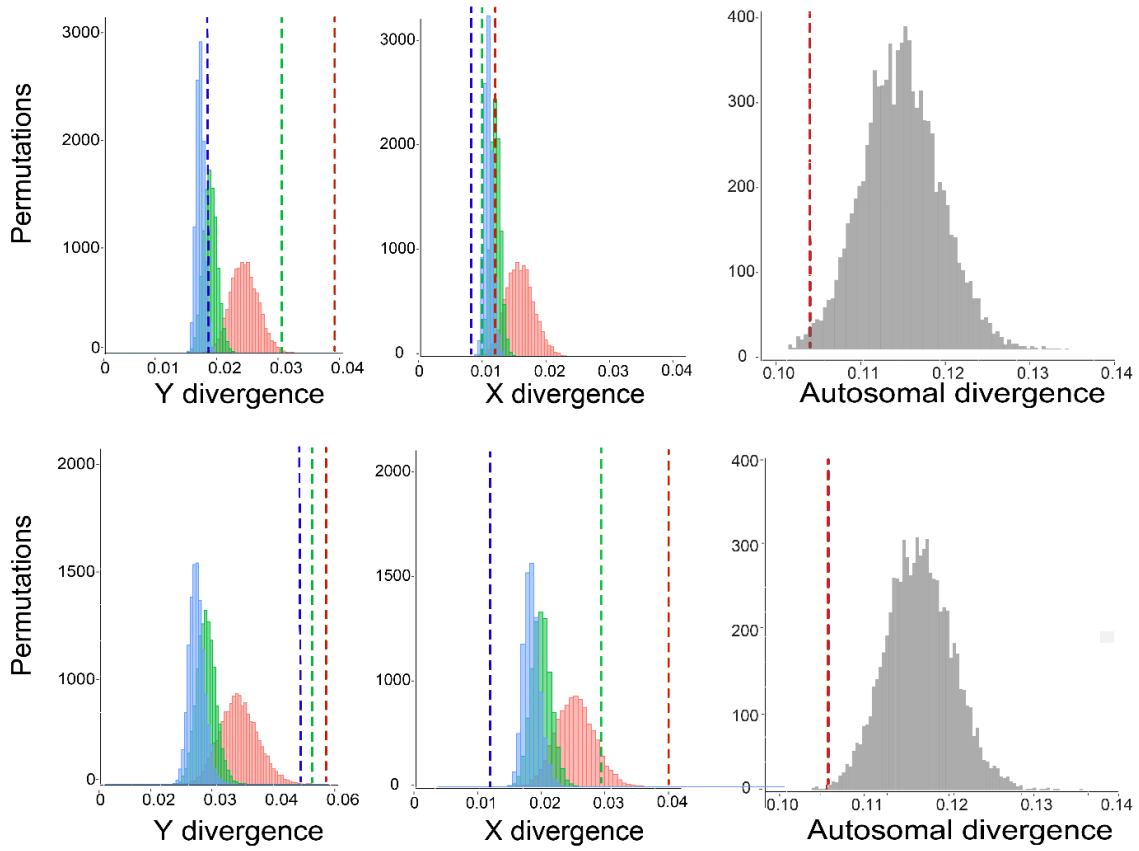
727 **Figures**

728



729

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



731

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

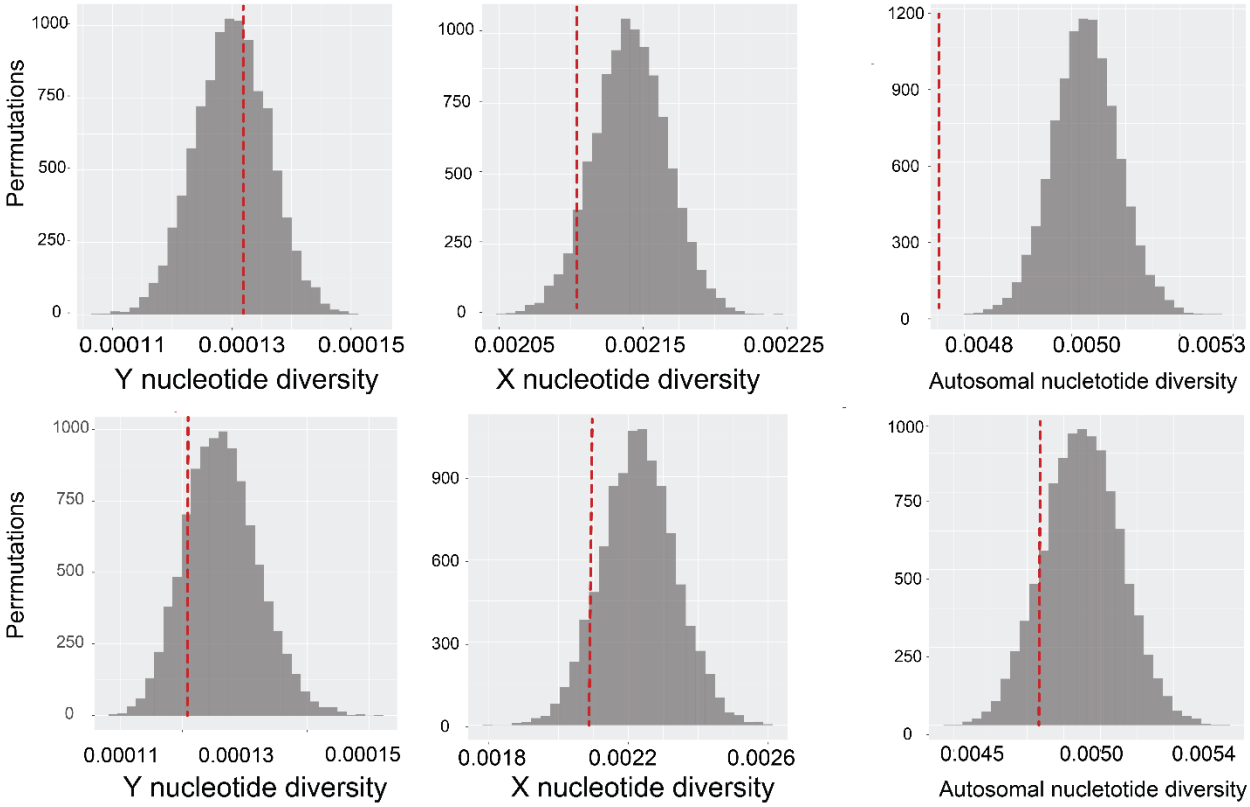
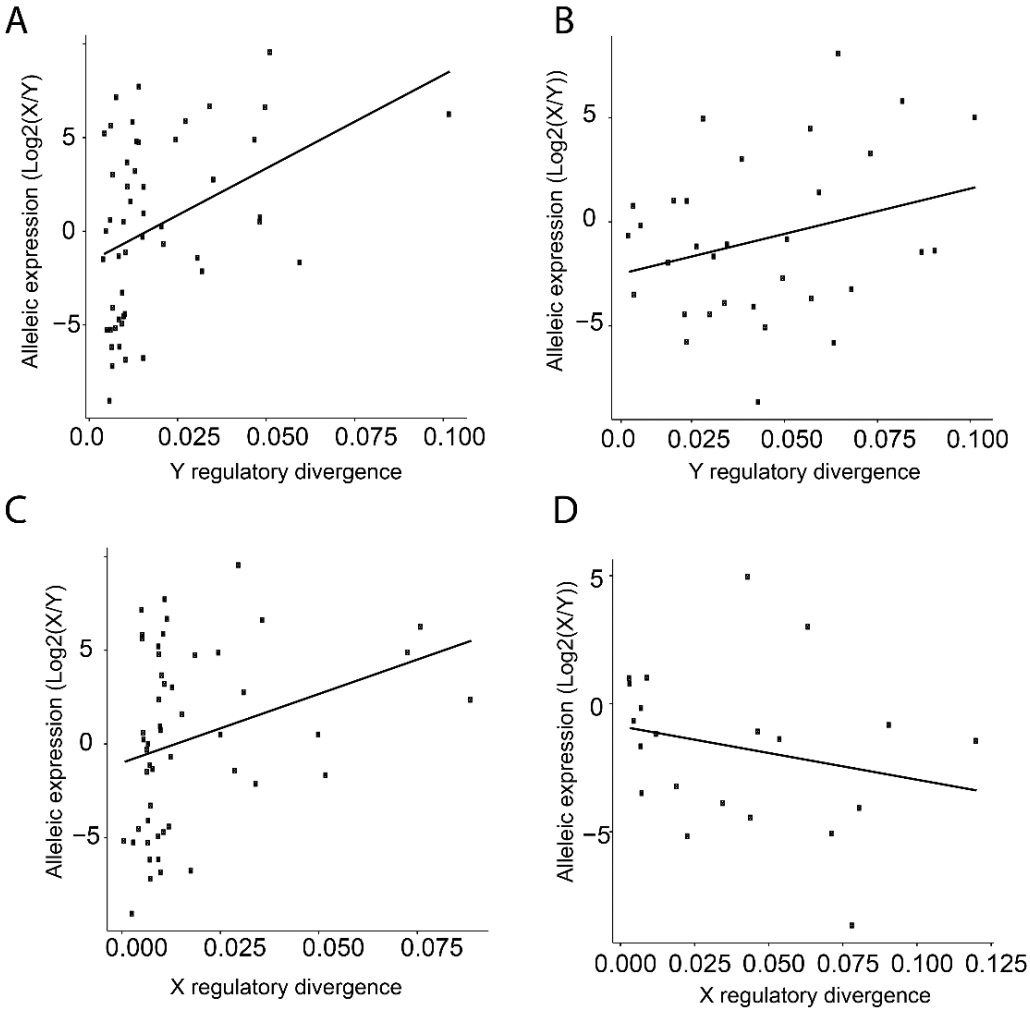


Figure 3. Regulatory nucleotide diversity (pi) compared to intergenic regions

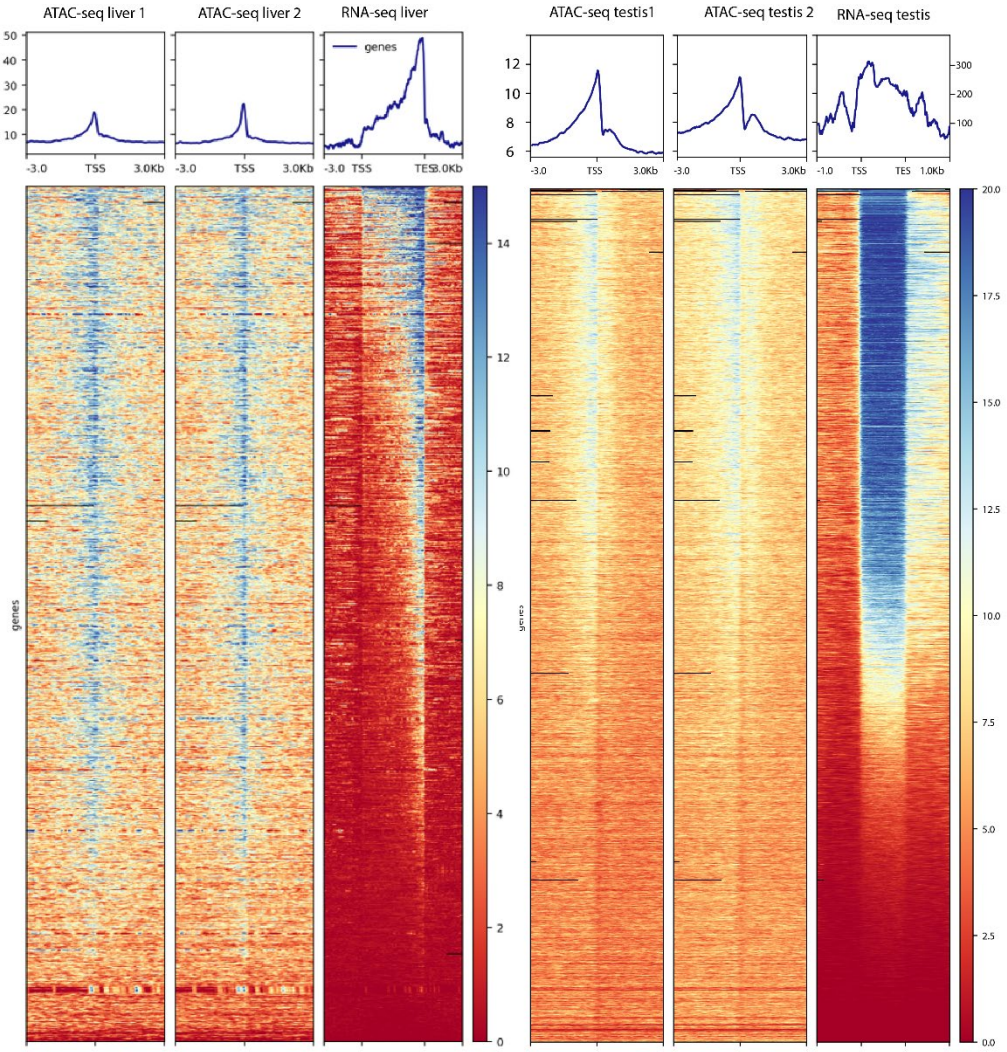


735

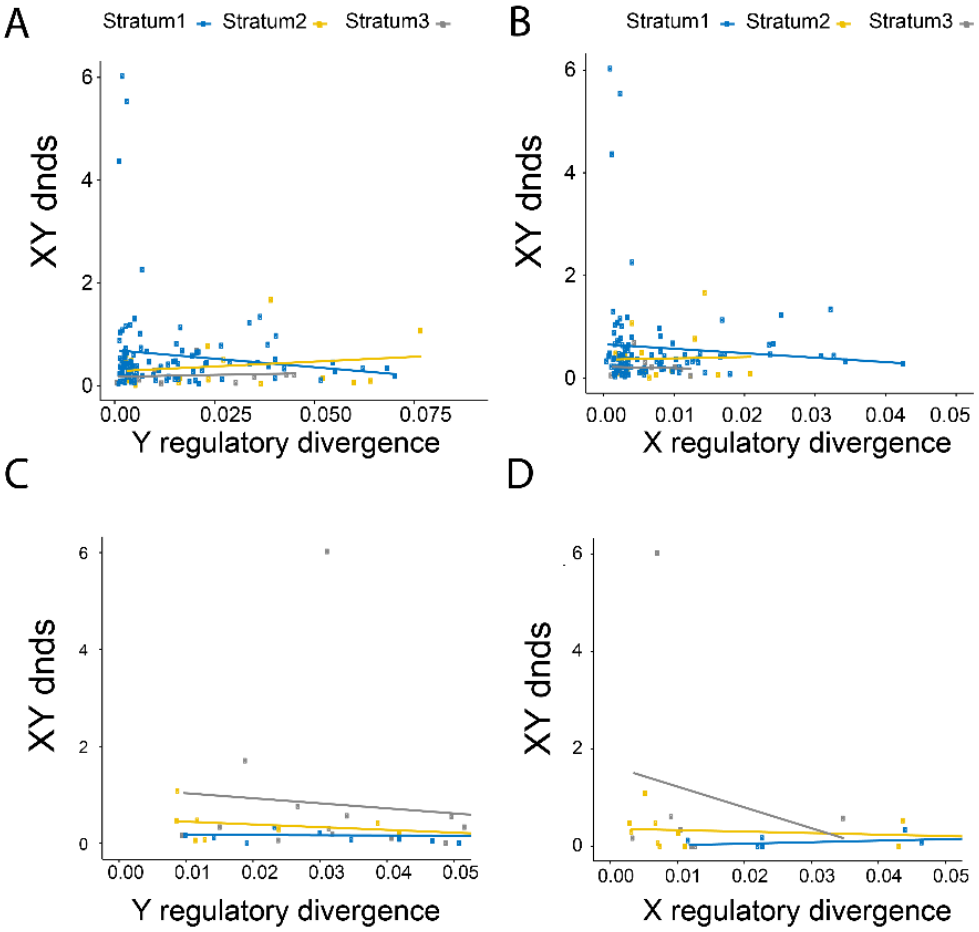
736 **Figure 4**. **Gametolog expression level compared to ACR divergence.**

737

Supplemental Figures



Supplemental Figure 1. High concordance of chromatin accessibility around transcription start site of expressed genes

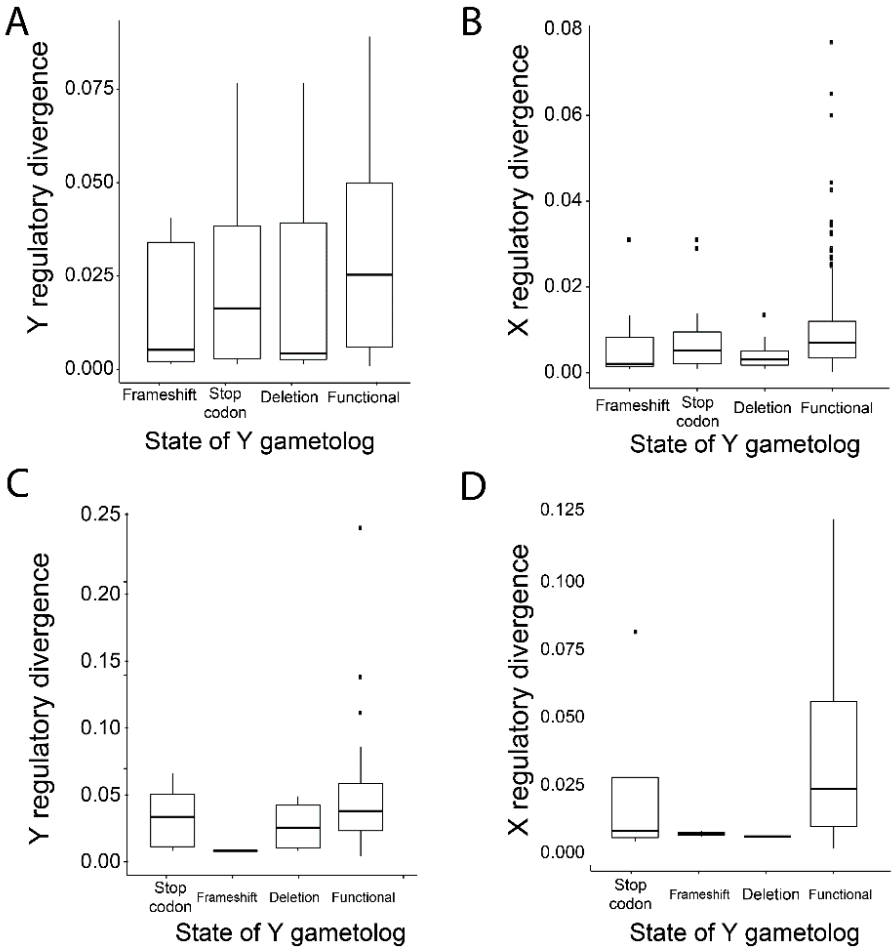


745

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

747

748



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