

Hill-Robertson Interference Reduced Genetic Diversity on a Young Plant Y-chromosome

Josh Hough^{*,†,1}, Wei Wang[†], Spencer C.H. Barrett[†] and Stephen I. Wright[†]

^{*}Department of Plant Sciences, University of California, Davis, [†]Department of Ecology and Evolutionary Biology, University of Toronto

ABSTRACT X and Y chromosomes differ in effective population size (N_e), rates of recombination, and exposure to natural selection, all of which can affect patterns of genetic diversity. On Y chromosomes with suppressed recombination, natural selection is expected to eliminate linked neutral variation and lower the N_e of Y compared to X chromosomes or autosomes. However, female-biased sex ratios and high variance in male reproductive success can also reduce Y-linked N_e , making it difficult to infer the causes of low Y-diversity. Here, we investigate the factors affecting levels of polymorphism during sex chromosome evolution in the dioecious plant *Rumex hastatus* (Polygonaceae). Strikingly, we find that neutral diversity for genes on the Y chromosome is on average 2.1% of the value for their X-linked homologues, corresponding to a chromosome-wide reduction of 93% compared to the standard neutral expectation. We demonstrate that the magnitude of this diversity loss is inconsistent with reduced male N_e caused by neutral processes. Instead, using forward simulations and estimates of the fitness effects of deleterious mutations, we show that Y chromosome diversity loss can be explained by purifying selection on a large number (≥ 800 kb) of genetically-linked sites. Our results are in agreement with theory on interference selection, and provide evidence that this can considerably reduce levels of polymorphism on a young plant Y chromosome. Given the relatively recent origin of *R. hastatus* sex chromosomes, our results imply that Y-chromosome degeneration in the early stages may be largely driven by such interference effects rather than by widespread gene silencing followed by neutral genetic drift.

KEYWORDS Nucleotide Diversity; Suppressed Recombination; Deleterious Mutations; Interference Selection

Introduction

Morphologically distinct sex chromosomes have evolved multiple times independently in both plants and animals (Westergaard 1958; Ohno 1967; Bull 1983; Charlesworth 1991, 2015). Despite clear biological differences between these kingdoms, X and Y chromosomes in both lineages have undergone similar genetic changes. In both groups, the loss of recombination between X and Y chromosomes is associated with an accumulation of deleterious mutations and a gradual loss of functional genes from the Y chromosome (Hough *et al.* 2014; Bergero *et al.* 2015; Bachtrog 2013), and in some species, such genetic deterioration has also led to the evolution of dosage compensation on the X chromosome (Charlesworth 1996b; Muyle

et al. 2012; Mank 2013; Papadopoulos *et al.* 2015). The independent evolution of these phenomena in such taxonomically distant groups as mammals and flowering plants suggests that general evolutionary mechanisms may be involved, but inferring the causes of molecular evolution and patterns of polymorphism in genomic regions that lack recombination is a long-standing challenge for both theoretical and experimental biologists (Charlesworth 1978; Feldman *et al.* 1980; Barton 1995; Charlesworth 1996b; Otto and Feldman 1997; Charlesworth and Charlesworth 2000; McVean and Charlesworth 2000a).

One fundamental difference between X and Y chromosomes is that there are 1/3 as many Y-linked gene copies as X-linked ones in a diploid population. Therefore, genes on the Y chromosome are expected to experience an effective population size (N_e) that is 1/4 that of autosomal genes, and 1/3 that of X-linked genes (assuming an equal number of reproducing females and males). The lowered N_e of the Y chromosome implies that the level of neutral polymorphism maintained at

equilibrium, proportional to the product of N_e and the neutral mutation rate, μ , should be lower for Y-linked genes than for their X-linked counterparts. In the absence of recombination, genes on the Y chromosome are also expected to be in strong linkage disequilibrium, making them vulnerable to diversity loss due to selection against strongly deleterious mutations (background selection; Charlesworth *et al.* 1993) and selective sweeps of strongly beneficial mutations (genetic hitchhiking; Maynard Smith and Haigh 1974). Furthermore, the build-up of genetic associations among selected mutations (fitness covariance) on the Y means that selection will act non-independently across the chromosome, such that selection at a focal site may "interfere" with selection at linked sites (Hill and Robertson 1966). A large body of work has now shown that such "selective interference" effects are expected to reduce both the efficacy of selection and the equilibrium level of neutral variability (Fisher 1930; Muller 1964; Hill and Robertson 1966; McVean and Charlesworth 2000b; Kaiser and Charlesworth 2009; Good *et al.* 2014), with the magnitude of the effect depending on the number and density of sites experiencing selection. These arguments all suggest that large non-recombining Y chromosomes with many linked selected sites should harbor a lower amount of neutral variability than predicted based on the number of Y chromosomes in a population (1/4 that of autosomes). However, the relative importance of neutral and selective factors for reducing chromosome-wide levels of diversity is not well understood.

In addition to reduced diversity arising from selection, in species with female-biased sex ratios or extensive male-male competition, high variance in male reproductive success is also expected to reduce the N_e experienced by genes on the Y chromosome (Caballero 1995; Charlesworth 2001; Laporte and Charlesworth 2002; Pool and Nielsen 2007; Ellegren 2009), suggesting that inferences about the effects of selection need to be distinguished from these neutral processes. Because variance in male reproductive success differentially affects the N_e of X, Y, and autosomal chromosomes (Kimura and Crow 1964; Nomura 2002), evidence for this can be obtained by comparing levels of silent site variability on X and Y chromosome, relative to values on autosomes. For example, variance in male reproductive success reduces male N_e , which reduces the Y/A diversity ratio relative to the neutral expectation (1/4), but also causes an increase in the X/A ratio. Based on such comparisons, studies in humans, for example, have suggested that the observation of an inflated X/A ratio can be explained by a historical excess of breeding females compared to males (Hammer *et al.* 2008; but see Bustamante and Ramachandran 2009; Hammer *et al.* 2010; Cotter *et al.* 2016).

Despite widespread interest in determining the evolutionary factors that affect neutral diversity on sex chromosomes (Ellegren 2011; Bachtrog 2013), we still know very little about how sex ratio variation or selective processes have affected levels of diversity on more recently evolved sex chromosomes. The time scales over which these processes are likely to be important are therefore not well understood. In humans, estimates of silent site diversity on the Y are considerably lower than predicted under neutral evolutionary models, and simulations suggest that observed levels of diversity are consistent with strong purifying selection (Wilson Sayres *et al.* 2014). However, given that human sex chromosomes evolved from autosomes ~200 million years ago (Lahn and Page 1999a; Ross *et al.* 2005), it is not clear whether purifying selection might have had a similarly strong

effect on Y chromosomes that evolved *do novo* from autosomes much more recently (e.g., within the last ~20 MYA in the case of dioecious plants; Charlesworth 2015).

Simulations of strong selection models (background selection and genetic hitchhiking) suggest that these processes may have the greatest effects during the earliest stages of sex chromosome evolution, before the Y has lost many of its genes (Bachtrog 2008). Moreover, theory suggests that even weak purifying selection, when aggregated over a large number of genetically-linked sites, can generate strong deviations from neutrality, whereas classic background selection theory breaks down in such cases (McVean and Charlesworth 2000b; Cameron and Kreitman 2002; Kaiser and Charlesworth 2009; Good *et al.* 2014). Given that a large number of selected sites are likely to be in linkage disequilibrium on a recently evolved Y chromosome, such "interference selection" *sensu* Good *et al.* 2014 is *a priori* likely to be a strong force affecting the evolution of young plant sex chromosomes. However, if there has been widespread gene silencing during the early stages of Y-chromosome evolution, as in *Drosophila albomicans*, for example (Zhou and Bachtrog 2012), then diversity loss might be expected to be lower on younger Y chromosomes. That is, if Y-linked gene silencing occurs early during sex chromosome evolution, then few sites may be under selection, and Y chromosome degeneration may be driven primarily by neutral genetic drift rather than inefficient selection. Understanding the relative importance of selection and neutral drift during Y chromosome evolution is therefore a major open issue.

To investigate the factors affecting nucleotide diversity in the early stages of sex chromosome evolution, we analyzed neutral polymorphism levels on X, Y, and autosomal chromosomes in the plant *Rumex hastatulus* (Polygonaceae). This species is a dioecious annual with heteromorphic X and Y chromosomes that originated ~15 MYA (Quesada del Bosque *et al.* 2011; Grabowska-Joachimiak *et al.* 2015; Navajas-Pérez *et al.* 2005), making Y chromosomes in this species over 100 million years younger than the highly degenerated Y chromosomes in mammals (Lahn and Page 1999b; Ross *et al.* 2005). *Rumex hastatulus* has also received particular attention because of the occurrence of an interesting polymorphism in sex chromosome system, in which both XY and XY₁Y₂ male genotypes occur in geographically distinct populations (so-called "chromosomal races"; Smith 1963). The derived XY₁Y₂ sex chromosome system in this species is thought to have originated through an X-autosome fusion, with the current XY system maintaining the ancestral chromosome complement (Smith 1964). Interestingly, despite the recent origin of sex chromosomes in both races, there is evidence that both the ancestral and neo-Y-chromosomes have undergone gene loss and functional deterioration (Hough *et al.* 2014). Here, to simplify our comparison of polymorphism levels on X, Y, and autosomes, and to ensure that our diversity estimates are not biased by non-equilibrium conditions arising from the X-autosome fusion, we focus only on the XY system with the ancestral chromosome complement.

Of particular relevance to our study, *R. hastatulus* populations have been found to exhibit female-biased reproductive sex ratios, with a mean sex ratio of ≈ 0.6 (Pickup and Barrett 2013). Indeed, the occurrence of female-biased sex ratios in dioecious flowering plants has been found to be more common in species with heteromorphic X and Y chromosomes than in those with monomorphic sex chromosomes (Field *et al.* 2013; Hough *et al.* 2013). In *R. hastatulus*, the availability of field measurements

173 of sex ratio variation thus provides an opportunity to test both
174 neutral and selective models of Y-chromosome diversity loss
175 during sex chromosome evolution.

176 Materials and Methods

177 Population Samples and Sex-Linked Genes

178 We analyzed sex-linked and autosomal genes identified from
179 Illumina RNA sequence data from 12 population samples (1
180 male and 1 female from each of 6 populations). Samples
181 were collected in 2010 from throughout the native range of
182 the Texas chromosome race *R. hastatus* (locations in Table
183 S1), and plants were grown in the glasshouse at the University
184 of Toronto from seeds collected from open-pollinated females.
185 We extracted RNA from leaf tissue using Spectrum Plant Total
186 RNA kits (Sigma-Aldrich). Isolation of mRNA and cDNA syn-
187 thesis was conducted according to standard Illumina RNAseq
188 procedures, and sequencing was conducted on two Illumina
189 HiSeq lanes with 150-bp end reads at the Genome Quebec In-
190 novation Center. Reads from these samples were mapped to
191 the *R. hastatus* reference transcriptome (Hough *et al.* 2014)
192 using the BurrowsWheeler Aligner (Li and Durbin 2010), fol-
193 lowed by Stampy (Lunter and Goodson 2011). We used Picard
194 tools (<http://picard.sourceforge.net>) to process mapping align-
195 ments for the Genome Analysis Toolkit variant calling software
196 (McKenna *et al.* 2010), and subsequently removed genes with
197 low coverage (<10x) and low Phred Quality Scores <20. The
198 RNAseq data from these population samples were previously
199 reported in Hough *et al.* (2014), where they were used to vali-
200 date the ascertainment of sex-linked genes identified through
201 segregation analysis, and raw sequences are available from the
202 GenBank Short Read Archive under accession no. SRP041588.
203 Here, to focus on sex-linked genes likely to be relatively older
204 and closer to equilibrium, we focused on the previously de-
205 scribed set of 460 X/Y genes for which a Y homolog was found
206 in both races (i.e., X/Y genes where the Y copy was inferred to
207 be on the Y₁ chromosome).

208 Autosomal Genes

209 In evaluating evidence for nucleotide diversity differences be-
210 tween X and Y chromosomes, it is important to distinguish be-
211 tween reduced Y-linked diversity, and the possibility that X-
212 linked diversity is elevated above the level predicted from a
213 neutral model. To do this, we normalized our sex-linked di-
214 versity estimates by autosomal diversity, and compared empiri-
215 cal X/A and Y/A nucleotide diversity ratios to those predicted
216 from neutral models and from simulations (described below).
217 Because the criteria for ascertaining autosomal loci in Hough
218 *et al.* (2014) were based on identifying four segregating SNPs
219 per locus, and since this set of genes is likely to be higher in di-
220 versity than the average autosomal gene, here we instead used
221 the larger set of all non-sex linked (putatively autosomal) genes
222 as our autosomal reference. We filtered genes in this set to re-
223 move any genes that may have been sex-linked but were not
224 identified as such by Hough et al.'s conservative ascertainment
225 criteria. In particular, we removed: (i) any genes in which there
226 was evidence for at least one SNP with a sex-linked segregation
227 pattern, (ii) any genes where SNPs showed fixed heterozygos-
228 ity in males and fixed homozygosity in females, (iii) genes with
229 less than 10X coverage or greater than 100X coverage from in-
230 dependently obtained genomic coverage data (Beaudry *et al.*,
231 in prep), to filter out duplicates or genes with highly repetitive
232 sequences, and (iv) any genes containing SNPs with large (>0.4)

233 allele frequency differences between males and females. Finally,
234 we removed genes with fewer than 100 filtered synonymous
235 sites to avoid biasing our results toward genes that may have
236 been particularly short due to assembly problems. This filter-
237 ing resulted in a final set of 12,356 autosomal genes.

238 Phasing X and Y alleles

239 To estimate polymorphism for X and Y sequences separately in
240 males, it is necessary to infer the phase of SNPs in sex-linked
241 transcripts. In previous work, phasing alleles on *R. hastatus*
242 sex chromosomes was achieved using segregation analysis
243 from a genetic cross. Here, to phase SNPs from population
244 samples where such segregation data was unavailable, we used
245 HAPCUT (Bansal and Batra 2008), a maximum-cut based algo-
246 rithm that reconstructs haplotypes using sequenced fragments
247 (Illumina read data) from the two homologous chromosomes
248 to output a list of phased haplotype blocks containing the poly-
249 morphic variants on each chromosome. Because the resulting
250 haplotype blocks produced by HAPCUT contained SNPs that
251 were phased relative to each other, but not designated to either
252 the X or Y chromosome, we assigned individual variants to X
253 or Y by independently identifying fixed X-Y differences within
254 each haplotype block (i.e., sites where all females were homozy-
255 gous, and all males were heterozygous). Identifying such fixed
256 differences within phased haplotype blocks enabled us to then
257 infer the correct phase (X or Y) of the polymorphisms from
258 HAPCUTs output. In particular, this was done by matching the
259 phase of fixed X-Y differences with neighboring polymorphic
260 sites: when a fixed X-Y difference occurred in the same phased
261 haplotype block as a polymorphic site, then the variants in that
262 block were assigned to either X or Y based on the known phase
263 of the fixed difference with which they were matched. SNPs
264 that were identified outside of phased blocks, or in blocks with-
265 out fixed X-Y differences, were recorded as missing data.
266 Finally, we filtered out SNPs with quality scores < 60, and those
267 within a distance of 10bp or less from indels. This filtering pro-
268 cedure resulted in alignments of X and Y sequences for 372 sex-
269 linked genes.

270 We further validated the results of HAPCUTs allele phasing
271 by comparing the accuracy of this method with the phasing-by-
272 segregation method that was conducted in Hough *et al.* (2014).
273 To do this, we first phased the sequence data from parents and
274 their progeny using HAPCUTs algorithm (using the same pa-
275 rameters as for the population data), and then identified cases
276 where SNPs were inferred on the Y chromosome by HAPCUT,
277 but where the true level of polymorphism, indicated by the
278 genetic cross, was zero. We identified 7% of sex-linked genes
279 that either had phasing errors or genotyping errors. This cor-
280 responds to a SNP error rate estimate of 1.7×10^{-4} . Note that
281 this rate is very low relative to population-based estimates of
282 polymorphism on the X and autosomes (Table 1), and therefore
283 should have minimal effects on our estimation of the X/A ratio.
284 However, because this rate is high relative to the expected level
285 of polymorphism on the Y chromosome, we conducted a fur-
286 ther filtering step for Y-linked SNPs and identified false positive
287 SNP calls arising from: (i) phasing errors caused by gene dup-
288 licates (more than two haplotypes), (ii) polymorphisms around
289 indels, and (iii) genotyping errors caused by low Y-expression.
290 This final filtering was conducted by manually checking each
291 individual putative polymorphism on the Y chromosome using
292 IGV (Robinson *et al.* 2011).

293 **Estimating nucleotide diversity on sex chromosomes and au-**
 294 **tosomes**

295 For each locus in our analysis, we calculated Watterson's (1975)
 296 estimator of the population parameter $\theta = 4N_e\mu$, where N_e is
 297 the diploid effective population size, and μ is the per-base mu-
 298 tation rate per-generation, using a modified version of the Perl
 299 program Polymorphurama (Bachtrog and Andolfatto 2006). To
 300 compare sex-linked and autosomal loci, we calculated the aver-
 301 age value of θ for each chromosome type, weighted by the num-
 302 ber of synonymous sites in each gene (Figure 2). We obtained
 303 95% confidence intervals for X/A and Y/A ratios by bootstrap-
 304 ping per gene using the BCa method (Efron and Tibshirani 1994)
 305 implemented in the Boot package in R (Canty and Ripley 2012),
 306 and calculating X/A and Y/A on each iteration for 20000 repli-
 307 cates each. Bootstrapping was conducted on the final filtered
 308 set of 173 sex-linked, and 12355 autosomal genes.

309 Note that the lack of recombination on the Y chromosome
 310 implies that statistical assumptions about independence across
 311 loci are violated, suggesting that the true uncertainty in the
 312 Y/A estimate may be wider than implied by bootstrapping. To
 313 address this, we also used a maximum likelihood approach
 314 (Wright and Charlesworth 2004) to independently estimate a
 315 credibility interval for the Y/A ratio (Figure S1). Because of
 316 the thousands of genes involved, a likelihood method incorpo-
 317 rating divergence to control for heterogeneity in mutation rate
 318 was not feasible, as this would require maximizing the like-
 319 lihood estimate of the mutation rate for each locus indepen-
 320 dently. However, previous analysis of divergence data does not
 321 suggest important chromosomal differences in synonymous di-
 322 versity contributing to diversity heterogeneity (see Results).
 323 Therefore, we assumed no heterogeneity in mutation rate, no
 324 recombination between Y-linked genes, and free recombination
 325 between autosomal loci. Our model thus had two parameters:
 326 θ_{aut} , our estimate of Watterson's theta for autosomal genes, and
 327 f , the ratio of effective population size of the Y chromosome to
 328 autosomes. We varied both parameters and evaluated the like-
 329 lihood for f from 0.001 to 1, and θ_{aut} per base pair from 0.001
 330 to 0.01. In particular, we evaluated the likelihood of θ_{aut} and
 331 f , given (1) the observed number of synonymous segregating
 332 sites for each i of n autosomal loci (S_i), (2) the number of segre-
 333 gating sites on the Y chromosome (S_y), and (3) the total number
 334 of sites per locus (L_i and L_y). The likelihood of (θ_{aut}, f) is given
 335 by:

$$336 L \propto P(S_Y, L_Y | f, \theta_{aut}) \prod_{i=1}^n P(S_i, L_i | \theta_{aut}) \quad (1)$$

336 **Neutral predictions and the effect of sex ratio bias on diversity**

337 To test whether our estimated levels of diversity on X, Y and au-
 338 tosomal chromosomes could be explained by neutral processes,
 339 including female-biased population sex ratios or high variance
 340 in male reproductive success, we compared our normalized
 341 X/A and Y/A diversity estimates to the corresponding neu-
 342 tral predictions for a given effective population size "sex ratio",
 343 $r = N_{ef}/(N_{ef} + N_{em})$. Because we were primarily interested in
 344 determining the parameter space within which a reduced male
 345 N_e could result in Y/A and X/A diversity ratios consistent with
 346 our data, we tested the fit of our estimates to predictions across
 347 the full range of r (0.1 to 0.9; we excluded all-female or all-male
 348 cases). For autosomal and sex-linked effective population sizes
 349 given by:

$$350 N_{e_A} = \frac{4N_{em}N_{ef}}{N_{em} + N_{ef}} \quad (2)$$

$$351 N_{ex} = \frac{9N_{em}N_{ef}}{4N_{em} + 2N_{ef}} \quad (3)$$

$$352 N_{ey} = \frac{N_{em}}{2} \quad (4)$$

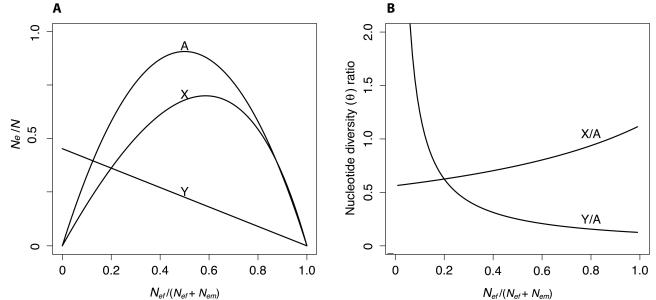
353 (Wright 1931), the corresponding expected X/A and Y/A ra-
 354 tios can be expressed as:

$$355 N_{ex} = \frac{9(-1+r)}{8(1-r)(-2+r)} \quad (5)$$

356 and

$$N_{ey} = \frac{1}{8r} \quad (6)$$

357 respectively. Note that when $N_{ef} = N_{em}$ ($r = 0.5$) then,
 358 $N_{ex}/N_{eA} = 0.75$, and $N_{ey}/N_{eA} = 0.25$ as in the standard neu-
 359 tral model (Wright 1931). With a female-biased ratio, however,
 360 the expected X/A ratio can become greater than 1, approaching
 361 1.125 in the limit as $r \rightarrow 1$ (Caballero 1995).



362 **Figure 1** **A.** Normalized effective population sizes at equilib-
 363 rium for genes on autosomes, X chromosomes, and Y chro-
 364 mosomes as a function of the sex ratio. **B.** The correspond-
 365 ing X/A and Y/A ratios of nucleotide diversity predicted at equi-
 366 librium, assuming equal neutral mutation rates for sex-linked
 367 and autosomal genes.

368 **Simulations of purifying selection**

369 To study the effects of purifying selection on expected levels
 370 of Y-chromosome diversity, we conducted forward-time sim-
 371 uations of haploid Y chromosomes using the software SFSCODE
 372 (Hernandez 2008). We first estimated the distribution of fitness
 373 effects of deleterious amino acid mutations from our polymor-
 374 phism data for X-linked genes using the method of Keightley
 375 and Eyre-Walker (2007), which fits a gamma distribution of neg-
 376 ative selection coefficients to the observed frequency distribu-
 377 tion of non-synonymous and synonymous polymorphisms. We
 378 then used this estimated gamma distribution to parameterize
 379 the simulations, initializing them with our estimated θ_{aut} , ad-
 380 justed to reflect the neutrally expected N_{ey} for a sex ratio of
 381 $r = 0.6$. To match our sample size and the number of syn-
 382 onymous sites sampled from our data (see Supporting Infor-
 383 mation), the simulations sampled 6 haploid chromosomes, and
 384 the genome sequence contained 45,331 bp of linked neutral se-
 385 quence from which we calculated silent site diversity, π_s . To
 386 examine the expected reduction in diversity as a function of the

number of selected sites (L), we ran simulations over a range of values of L , up to a maximum of 5×10^6 (Figure 3). To obtain an estimate of L , we calculated the approximate likelihood of our observed data based on the proportion of simulations in which synonymous diversity, π_s , was less than 0.0001 from our empirical estimate (Figure 3B).

Results and Discussion

Extensive loss of Y-chromosome diversity

Our analysis of polymorphism levels across the genome of the dioecious plant *R. hastatus* revealed a widespread loss of neutral diversity from the Y chromosome. In particular, neutral diversity for Y-linked genes was on average ~2.1% of the value for their homologues on the X chromosome. Taking 1/4 of the mean autosomal diversity as the equilibrium expectation for θ_Y under neutrality, this corresponds to a chromosome-wide reduction of 93% relative to the standard neutral model (Wright 1931). Note that by normalizing X and Y diversity estimates by autosomal diversity levels, our results indicate that the difference between X and Y homologues is not due to an elevation of diversity on the X chromosome, but a Y-specific reduction. Interestingly, our results also suggest an elevation of X-linked diversity ($X/A = 0.85$) compared to the neutral prediction ($X/A = 0.75$), though the confidence intervals on this estimate are wide, ranging from 0.74 to 0.95 (Figure 2). Given the empirically-estimated sex ratio of $r = 0.6$ in *R. hastatus* (Pickup and Barrett 2013), however, the X/A elevation we observed is not unexpected; the neutrally-predicted X/A ratio for a sex ratio of 0.6 is ≈ 0.8 (Figure 2; Equation 4). Although our estimates of diversity have not been normalized by divergence, previous work has shown that the average synonymous substitution rate, between *R. hastatus* and the non-dioecious out-group *R. bucephalophorus* is comparable for sex-linked (0.2016) and autosomal genes (0.219), and we found no evidence for significant differences in substitution rate between Y and X chromosomes (Hough et al. 2014). It is therefore unlikely that our results are caused by mutation rate differences between sex-linked and autosomal genes.

Table 1 Observed and neutrally-expected silent site diversity on sex chromosomes and autosomes in *R. hastatus*

chr	Observed		Expected	
	θ	$/\theta_{aut}$	θ	$/\theta_{aut}$
A	0.0055	1	0.0055	1
X	0.0047	0.85	0.0041	0.75
Y	10^{-4}	0.018	0.0014	0.25

Female Biased Sex Ratios and Variance in Male Reproductive Success

The occurrence of female-biased sex ratios in *R. hastatus* is expected to lower Y diversity through a reduction in male N_e and thus a reduction in the N_e of the Y chromosome. Male N_e could be further reduced by high variance in male reproductive success, which is expected in annual wind-pollinated plants such as *R. hastatus* that commonly exhibit extensive phenotypic plasticity in plant size and flower production. Given that male

plants in this species produce large amounts of pollen, and female flowers are uniovulate, there may indeed be strong competition among males to fertilize ovules. Such competition should cause a proportional increase in X-linked diversity to a level that is close to (or even higher than) levels of autosomal diversity (Caballero 1995), while simultaneously reducing Y-linked diversity. In common with most flowering plants, we do not have marker-based estimates of variance in male reproductive success in *R. hastatus*. We therefore tested whether an overall reduction in male N_e , arising either from high variance in male reproductive offspring number and/or a female biased population sex ratio, could explain our observed Y/A and X/A ratios, by comparing our data to neutral predictions across the full range of values for the N_e sex ratio, r .

As shown in Figure 2, the expected reduction in Y/A as a function of r is substantially higher than our observed Y/A diversity ratio. Indeed, the lower limit for the Y/A ratio in a neutral model is 1/8, such that even in the extreme case where $r = 0.9$, the expectation for Y/A is ≈ 0.14 , which is substantially higher than our observed estimate ($Y/A = 0.018$) (Table 1; Figure 2). Moreover, such large reductions in N_{e_m} would also predict levels of X/A diversity that are significantly larger than what we observed (Figure 2). Our results therefore indicate that, although reduced male N_e arising from sex-biased demography is expected to contribute to reduced Y chromosome polymorphism, it is insufficient to explain the Y/A reduction we observed.

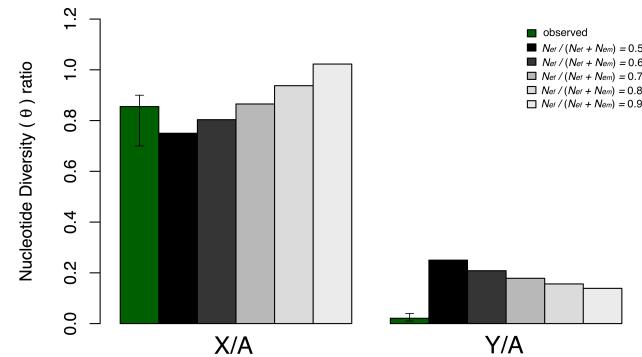


Figure 2 Observed and predicted X/A and Y/A ratios of neutral diversity. Predictions are shown for increasing values of the effective population size sex ratio, $N_{e_f}/(N_{e_m} + N_{e_f})$. Observed estimates of X/A and Y/A were calculated as the average θ across genes, weighted by the number of synonymous sites per gene. Confidence intervals were calculated by bootstrapping (20000 replicates) using the BCa method (Efron 1987) implemented in the Boot package in R (Canty and Ripley 2012), and using maximum likelihood for the Y-chromosome (see Methods).

Purifying selection

Selection against strongly deleterious mutations is expected to be an important factor reducing genetic variability in regions with low recombination (Charlesworth et al. 1993; Charlesworth 1996a). The effects of such background selec-

tion (BGS) have been well studied theoretically (Charlesworth *et al.* 1997; Nordborg *et al.* 1996; Kim and Stephan 2000) and the theory has been used to explain patterns of genetic diversity across genomes (Comeron 2014) and plays a central role in explanations for diversity loss on Y chromosomes in mammals (Wilson Sayres *et al.* 2014) and *Drosophila* (McAllister and Charlesworth 1999; Charlesworth 1996b). Under background selection theory for a Y chromosome, the predicted loss of diversity is modeled as a reduction in the equilibrium N_e such that $\pi \approx (\frac{Y}{A})4N_e f_0 \mu$, where $f_0 = e^{-\frac{U}{sh}}$, U is the mutation rate to strongly deleterious variants on the Y, and s and h are selection and dominance coefficients (Hudson and Kaplan 1995). This yields the expected value of π under the model: $E[\pi] = \pi_0 e^{-\frac{U}{sh}}$, where π_0 is the expected diversity in the absence of BGS, $\pi_0 = (\frac{Y}{A})4N_e \mu$, and Y/A is the expected ratio of the effective population size of the Y relative to autosomes. Here, assuming that the proportion of males is 0.4, and therefore that Y/A is 0.21 of the total N_e estimated from autosomes as a result of the female biased sex ratio in *R. hastatulus* (Pickup and Barrett 2013), and using $s=0.003$ as estimated from the distribution of fitness effects of deleterious mutations, the corresponding expected diversity on the Y chromosome under BGS is $E[\pi] = 1.5 \times 10^{-5}$ for a haploid Y chromosome mutation rate of 7×10^{-9} per nucleotide per generation and 1 MB of selected sites. Although there is considerable uncertainty in these parameters, the calculation yields an expected value of diversity that is 70% lower than what we observed (Table 1; Figure 2; Figure 4).

Importantly, BGS assumes independence among sites, and the theory breaks down when many linked sites are subject to selection at the same time. This is because high linkage disequilibrium causes the behavior of selected alleles interfere with the action of selection at linked sites (Good *et al.* 2014; Kaiser and Charlesworth 2009). As there are no analytical formulae for predicting the outcome of selection on many linked alleles experiencing selection and drift, however, simulations are fundamental to understanding such selective interference in realistic situations. We therefore conducted forward simulations of purifying selection using our estimated parameters for the distribution of fitness effects of deleterious mutations, and tested whether interference selection could result in a level of Y diversity at neutral sites similar to the level we observed. Our maximum likelihood estimates of the number of selected sites, L , from our simulations, suggest that the number of sites subject to selection required to explain the diversity loss from the *R. hastatulus* Y chromosome is ($\approx 800kb$) and may be as large as 2.5 MB (Figure 3). Thus, these results are consistent with the hypothesis that the early stages of Y degeneration are characterized by the persistence of a large number of sites subject to purifying selection.

While clearly approximate, we can ask how these estimates compare to how many selected sites we might expect to be present on the Y chromosome. Our previous estimates based on cytological measurements suggested that there may be roughly 5,600 genes on the X chromosome of the XY race (Hough *et al.* 2014). We also estimated that approximately 28 percent of genes had degenerated, leaving us with an estimated 1568 genes remaining on the Y chromosome. Using the average number of codons in genes of *Arabidopsis thaliana* (426) (Wortman *et al.* 2003), and assuming 2/3 of codons are under selection, this would imply 1.3 MB of selected sites, which is well within the range of our uncertainty (Figure 3). Thus, although our estimates are broad, these results are consistent with the retention

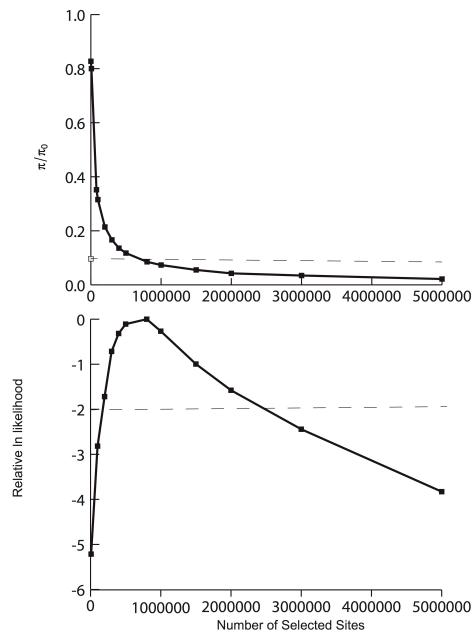


Figure 3 Results from forward simulations of purifying selection. Top: estimated silent site diversity π relative to neutrality (π_0) as a function of the number of selected sites. The dashed line corresponds to our estimate of diversity on the Y chromosome (θ_Y). Bottom: the relative likelihood curve for the number of selected sites, with the dashed line reflecting the approximate 95% credibility interval.

of a large proportion of sites subject to purifying selection on the Y chromosome, as expected in the early stages of degeneration.

Although our analyses of purifying selection models are consistent with our observed reduction in Y-linked diversity, it is also possible that selective sweeps have contributed. We have not formally excluded this possibility. However, if patterns of molecular evolution on the Y are driven primarily by interference selection, as our results suggest, then a further implication is that beneficial mutations should suffer a reduction in their probability of fixation, limiting the scope for sweeps to drive down diversity. Nonetheless, there is abundant evidence that soft sweeps are important determinants of patterns of nucleotide diversity (Messer and Petrov 2013), and it is possible that they have contributed to the extremely low levels of Y-linked variability that we observed.

Conclusions

The non-recombining region of a Y chromosome produces evolutionary dynamics that are similar to a haploid asexual population, with the lack of recombination resulting in build up of genetic associations among selected sites, a loss in the efficiency of selection, and an increase in population fitness variance

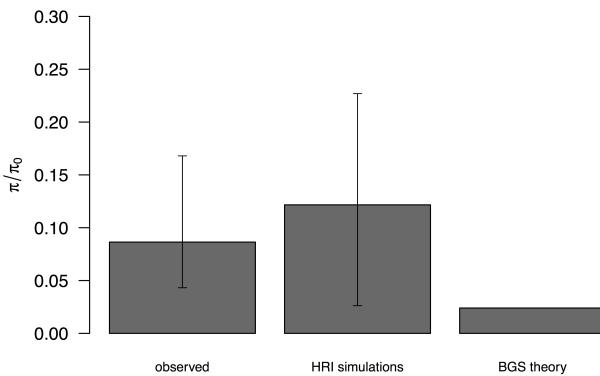


Figure 4 Estimated levels of Y-linked nucleotide diversity compared to estimates obtained from simulations of purifying selection among linked selected sites (HRI simulations) and analytical predictions from background selection theory. Silent site diversity is reported relative to the neutral expectation (π_0), where $\pi_0 = (\frac{Y}{A})4N_e\mu$, and Y/A is the expected ratio of the effective population size of the Y relative to autosomes for a sex ratio of $r = 0.6$, as estimated in *R. hastatus* populations. The credibility interval for our observed Y diversity was obtained by maximum likelihood (see methods), and error bars for HRI simulations correspond to 1.5x the interquartile range for π_s estimates, approximately equal to the 95% confidence interval for the median (Chambers 1983).

(Fisher 1930; Muller 1964; Hill and Robertson 1966; McVean and Charlesworth 2000b; Kaiser and Charlesworth 2009; Good *et al.* 2014). Our study of neutral diversity levels on the relatively young *R. hastatus* sex chromosomes provides clear evidence for an extensive loss of neutral diversity on the Y chromosome. Whereas neutral models of sex-biased demography were unable to explain the magnitude of diversity loss, forward population genetic simulations suggest that the extensive loss of diversity is most likely due to purifying selection. However, while standard background selection theory over-predicted the loss of diversity on the Y chromosome by $\approx 70\%$, simulations of purifying selection acting over a large number of linked selected sites resulted in a level of Y-diversity that was similar to what we observed. These results are consistent with theory on interference selection, and suggest that the effects of interference is likely an important force in the evolution of young Y chromosomes. Moreover, our results imply that when a large number of linked sites are subject to purifying selection, Y-chromosome degeneration in the early stages may largely be driven by the effects of interference rather primarily by widespread silencing of Y-linked genes and their subsequent degeneration through neutral drift.

Acknowledgments

We thank Felix Baudry for providing genomic coverage data for quality filtering. This research was funded by Discovery grants to SCHB and SIW from the Natural Sciences and Engineering Research Council of Canada, and by an OGS scholarship to JH.

Literature Cited

- Bachtrog, D., 2008 The temporal dynamics of processes underlying y-chromosome degeneration. *Genetics* **179**: 1513–1525.
- Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of y-chromosome degeneration. *Nature Reviews Genetics* **14**: 113–124.
- Bachtrog, D. and P. Andolfatto, 2006 Selection, recombination and demographic history in *Drosophila miranda*. *Genetics* **174**: 2045–2059.
- Bansal, V. and V. Bafna, 2008 Hapcut: an efficient and accurate algorithm for the haplotype assembly problem. *Bioinformatics* **24**: i153–i159.
- Barton, N., 1995 A general model for the evolution of recombination. *Genetical research* **65**: 123–144.
- Bergero, R., S. Qiu, and D. Charlesworth, 2015 Gene loss from a plant sex chromosome system. *Current Biology* **25**: 1234–1240.
- Bull, J., 1983 *Evolution of sex determining mechanisms*. The Benjamin/Cummings Publishing Company, Inc.
- Bustamante, C. D. and S. Ramachandran, 2009 Evaluating signatures of sex-specific processes in the human genome. *Nature genetics* **41**: 8–10.
- Caballero, A., 1995 On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**: 1007–1011.
- Canty, A. and B. Ripley, 2012 boot: Bootstrap r (s-plus) functions. R package version 1.
- Chambers, J. M., 1983 *Graphical methods for data analysis*.
- Charlesworth, B., 1978 Model for evolution of y-chromosome and dosage compensation. *Proceedings of the National Academy of Sciences* **75**: 5618–5622.
- Charlesworth, B., 1991 The evolution of sex chromosomes. *Science* **251**: 1030–1033.
- Charlesworth, B., 1996a Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical research* **68**: 131–149.
- Charlesworth, B., 1996b The evolution of chromosomal sex determination and dosage compensation. *Current Biology* **6**: 149–162.
- Charlesworth, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. *Genetical research* **77**: 153–166.
- Charlesworth, B. and D. Charlesworth, 2000 The degeneration of y-chromosome. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**: 1563–1572.
- Charlesworth, B., M. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical research* **70**: 155–174.
- Charlesworth, D., 2015 Plant contributions to our understanding of sex chromosome evolution. *New Phytologist* **208**: 52–65.
- Comeron, J. M., 2014 Background selection as baseline for nucleotide variation across the drosophila genome. *PLoS Genet* **10**: e1004434.
- Comeron, J. M. and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**: 389–410.
- Cotter, D. J., S. M. Brotman, and M. A. W. Sayres, 2016 Genetic

- 627 diversity on the human x chromosome does not support a 689
 628 strict pseudoautosomal boundary. *Genetics* **203**: 485–492.
 629 Efron, B., 1987 Better bootstrap confidence intervals. *Journal of 690*
 630 the American statistical Association **82**: 171–185.
 631 Efron, B. and R. J. Tibshirani, 1994 *An introduction to the bootstrap*.
 632 CRC press.
 633 Ellegren, H., 2009 The different levels of genetic diversity in 691
 634 sex chromosomes and autosomes. *Trends in Genetics* **25**: 278–
 635 284.
 636 Ellegren, H., 2011 Sex-chromosome evolution: recent progress 692
 637 and the influence of male and female heterogamety. *Nature 693*
 638 *Reviews Genetics* **12**: 157–166.
 639 Feldman, M. W., F. B. Christiansen, and L. D. Brooks, 1980 694
 640 Evolution of recombination in a constant environment. *Proceedings 695*
 641 of the National Academy of Sciences **77**: 4838–4841.
 642 Field, D. L., M. Pickup, and S. C. Barrett, 2013 Comparative 696
 643 analyses of sex-ratio variation in dioecious flowering plants. 697
 644 *Evolution* **67**: 661–672.
 645 Fisher, R. A., 1930 *The genetical theory of natural selection: a 698*
 646 complete variorum edition. Oxford University Press.
 647 Good, B. H., A. M. Walczak, R. A. Neher, and M. M. Desai, 2014 699
 648 Genetic diversity in the interference selection limit. *PLoS 700*
 649 *Genet* **10**: e1004222.
 650 Grabowska-Joachimiak, A., A. Kula, T. Książczyk, J. Chojnicka, 701
 651 E. Sliwinska, and A. J. Joachimiak, 2015 Chromosome 712
 652 landmarks and autosome-sex chromosome translocations in 713
 653 *rumex hastatus*, a plant with xx/xyly sex chromosome 714
 654 system. *Chromosome Research* **23**: 187–197.
 655 Hammer, M. F., F. L. Mendez, M. P. Cox, A. E. Woerner, and 715
 656 J. D. Wall, 2008 Sex-biased evolutionary forces shape genomic 716
 657 patterns of human diversity. *PLoS Genet* **4**: e1000202.
 658 Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, M. P. 717
 659 Cox, and J. D. Wall, 2010 The ratio of human x chromosome to 718
 660 autosome diversity is positively correlated with genetic dis- 719
 661 tance from genes. *Nature genetics* **42**: 830–831.
 662 Hernandez, R. D., 2008 A flexible forward simulator for popula- 720
 663 tions subject to selection and demography. *Bioinformatics* **24**: 721
 664 2786–2787.
 665 Hill, W. G. and A. Robertson, 1966 The effect of linkage on limits 722
 666 to artificial selection. *Genetical research* **8**: 269–294.
 667 Hough, J., J. D. Hollister, W. Wang, S. C. Barrett, and S. I. Wright, 723
 668 2014 Genetic degeneration of old and young y-chromosomes 724
 669 in the flowering plant *rumex hastatus*. *Proceedings of the 725*
 670 National Academy of Sciences **111**: 7713–7718.
 671 Hough, J., S. Immler, S. C. Barrett, and S. P. Otto, 2013 Evolu- 726
 672 tionarily stable sex ratios and mutation load. *Evolution* **67**: 727
 673 1915–1925.
 674 Hudson, R. R. and N. L. Kaplan, 1995 Deleterious background 728
 675 selection with recombination. *Genetics* **141**: 1605–1617.
 676 Kaiser, V. B. and B. Charlesworth, 2009 The effects of deleterious 729
 677 mutations on evolution in non-recombining genomes. *Trends 730*
 678 in Genetics **25**: 9–12.
 679 Keightley, P. D. and A. Eyre-Walker, 2007 Joint inference of the 731
 680 distribution of fitness effects of deleterious mutations and 732
 681 population demography based on nucleotide polymorphism 733
 682 frequencies. *Genetics* **177**: 2251–2261.
 683 Kim, Y. and W. Stephan, 2000 Joint effects of genetic hitchhiking 734
 684 and background selection on neutral variation. *Genetics* **155**: 735
 685 1415–1427.
 686 Kimura, M. and J. F. Crow, 1964 The number of alleles that can 736
 687 be maintained in a finite population. *Genetics* **49**: 725–738.
 688 Lahn, B. T. and D. C. Page, 1999a Four evolutionary strata on 737
 689 the human x chromosome. *Science* **286**: 964–967.
 690 Lahn, B. T. and D. C. Page, 1999b Four evolutionary strata on 691
 692 the human x chromosome. *Science* **286**: 964–967.
 693 Laporte, V. and B. Charlesworth, 2002 Effective population size 694
 694 and population subdivision in demographically structured 695
 695 populations. *Genetics* **162**: 501–519.
 696 Li, H. and R. Durbin, 2010 Fast and accurate long-read align- 697
 697 ment with burrows–wheeler transform. *Bioinformatics* **26**: 698
 698 589–595.
 699 Lunter, G. and M. Goodson, 2011 Stampy: a statistical algo- 700
 700 rithm for sensitive and fast mapping of illumina sequence 701
 701 reads. *Genome research* **21**: 936–939.
 702 Mank, J. E., 2013 Sex chromosome dosage compensation: defi- 703
 703 nitely not for everyone. *Trends in genetics* **29**: 677–683.
 704 Maynard Smith, J. and J. Haigh, 1974 The hitch-hiking effect of 705
 705 a favourable gene. *Genetical research* **23**: 23–35.
 706 McAllister, B. F. and B. Charlesworth, 1999 Reduced sequence 707
 707 variability on the neo-y chromosome of *Drosophila americana*. 708
 708 *Genetics* **153**: 221–233.
 709 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibul- 710
 710 skis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, 711
 711 M. Daly, et al., 2010 The genome analysis toolkit: a map- 712
 712 reduce framework for analyzing next-generation dna sequenc- 713
 713 ing data. *Genome research* **20**: 1297–1303.
 714 McVean, G. A. and B. Charlesworth, 2000a The effects of hill- 715
 715 robertson interference between weakly selected mutations on 716
 716 patterns of molecular evolution and variation. *Genetics* **155**: 717
 717 929–944.
 718 McVean, G. A. and B. Charlesworth, 2000b The effects of hill- 719
 719 robertson interference between weakly selected mutations on 720
 720 patterns of molecular evolution and variation. *Genetics* **155**: 721
 721 929–944.
 722 Messer, P. W. and D. A. Petrov, 2013 Population genomics of 723
 723 rapid adaptation by soft selective sweeps. *Trends in ecology & 724
 724 evolution* **28**: 659–669.
 725 Muller, H. J., 1964 The relation of recombination to mutational 726
 726 advance. *Mutation Research/Fundamental and Molecular 727
 727 Mechanisms of Mutagenesis* **1**: 2–9.
 728 Muyle, A., N. Zemp, C. Deschamps, S. Mousset, A. Widmer, 729
 729 and G. A. Marais, 2012 Rapid de novo evolution of x chromo- 730
 730 some dosage compensation in *Silene latifolia*, a plant with 731
 731 young sex chromosomes. *PLoS Biol* **10**: e1001308.
 732 Navajas-Pérez, R., R. de la Herrán, G. L. González, M. Jamilena, 733
 733 R. Lozano, C. R. Rejón, M. R. Rejón, and M. A. Garrido- 734
 734 Ramos, 2005 The evolution of reproductive systems and 735
 735 sex-determining mechanisms within *rumex* (polygonaceae) 736
 736 inferred from nuclear and chloroplastidial sequence data. 737
 737 *Molecular Biology and Evolution* **22**: 1929–1939.
 738 Nomura, T., 2002 Effective size of populations with unequal 739
 739 sex ratio and variation in mating success. *Journal of Animal 740
 740 Breeding and Genetics* **119**: 297–310.
 741 Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The 742
 742 effect of recombination on background selection. *Genetical 743
 743 research* **67**: 159–174.
 744 Ohno, S., 1967 *Sex chromosomes and sex-linked genes*, volume 1. 745
 745 Springer-Verlag.
 746 Otto, S. P. and M. W. Feldman, 1997 Deleterious mutations, vari- 747
 747 able epistatic interactions, and the evolution of recombina- 748
 748 tion. *Theoretical population biology* **51**: 134–147.
 749 Papadopoulos, A. S., M. Chester, K. Ridout, and D. A. Filatov, 750
 750 2015 Rapid y degeneration and dosage compensation in plant
 751 sex chromosomes. *Proceedings of the National Academy of*

- 751 Sciences **112**: 13021–13026.
- 752 Pickup, M. and S. C. Barrett, 2013 The influence of demogra-
753 phy and local mating environment on sex ratios in a wind-
754 pollinated dioecious plant. Ecology and evolution **3**: 629–639.
- 755 Pool, J. E. and R. Nielsen, 2007 Population size changes reshape
756 genomic patterns of diversity. Evolution **61**: 3001–3006.
- 757 Quesada del Bosque, M., R. Navajas-Pérez, J. Panero,
758 A. Fernández-González, and M. Garrido-Ramos, 2011 A satel-
759 lite dna evolutionary analysis in the north american endemic
760 dioecious plant rumex hastatus (polygonaceae). Genome
761 **54**: 253–260.
- 762 Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman,
763 E. S. Lander, G. Getz, and J. P. Mesirov, 2011 Integrative ge-
764 nomics viewer. Nature biotechnology **29**: 24–26.
- 765 Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay,
766 D. Muzny, M. Platzer, G. R. Howell, C. Burrows, C. P. Bird,
767 *et al.*, 2005 The dna sequence of the human x chromosome.
768 Nature **434**: 325–337.
- 769 Smith, B. W., 1963 The mechanism of sex determination in
770 rumex hastatus. Genetics **48**: 1265.
- 771 Smith, B. W., 1964 The evolving karyotype of rumex hastatus.
772 Evolution pp. 93–104.
- 773 Westergaard, M., 1958 The mechanism of sex determination in
774 dioecious flowering plants. Advances in genetics **9**: 217–281.
- 775 Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen,
776 2014 Natural selection reduced diversity on human y-
777 chromosomes. PLoS Genet **10**: e1004064.
- 778 Wortman, J. R., B. J. Haas, L. I. Hannick, R. K. Smith, R. Maiti,
779 C. M. Ronning, A. P. Chan, C. Yu, M. Ayele, C. A. Whitelaw,
780 *et al.*, 2003 Annotation of the arabidopsis genome. Plant Phys-
781 iology **132**: 461–468.
- 782 Wright, S., 1931 Evolution in mendelian populations. Genetics
783 **16**: 97–159.
- 784 Wright, S. I. and B. Charlesworth, 2004 The hka test revisited.
785 Genetics **168**: 1071–1076.
- 786 Zhou, Q. and D. Bachtrog, 2012 Chromosome-wide gene silenc-
787 ing initiates y degeneration in drosophila. Current Biology **22**:
788 522–525.