Nucleotide Diversity Loss on a Plant Y Chromosome Following Recent Recombination Suppression

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 significantly affect levels of genetic variability within the genome. On Y chromosomes with suppressed recombination, selection can remove genetically-linked neutral variation, resulting in reduced N_e and the level of polymorphism on Y compared to X chromosomes or autosomes. However, non-selective factors including female biased sex ratios and high variance in male fitness can also reduce Y-linked N_e, making it difficult to infer the causes of low Y-chromosome diversity. Here, we investigate the factors affecting X- and Y-linked diversity during plant sex chromosome evolution in *Rumex hastatulus* (Polygonaceae). Strikingly, we find that average neutral polymorphism levels for genes on the Y chromosome are ~2.1% the value of their homologues on the X chromosome, indicating a significant departure from neutrality and a chromosome-wide reduction of ~93%. We demonstrate that the magnitude of this diversity loss is inconsistent with neutral models that account for the occurrence of female-biased sex ratios, or by reduced male N_e caused by high variance in male fitness. Rather, using forward simulations and estimates of the distribution of fitness effects of deleterious mutations, we show that neutral polymorphism on the Y can be explained by the effects of linked purifying selection, although selective sweeps may have also contributed. Given the recent origin of *R. hastatulus* sex chromosomes, our study suggests that the early stages of Y chromosome degeneration are characterized by strong selective interference among many linked sites rather than gene silencing followed by neutral genetic drift.

KEYWORDS Sex Chromosome Evolution; Nucleotide Diversity; Recombination; Deleterious Mutations

Introduction

Morphologically distinct sex chromosomes have evolved multiple times independently in both plant and animal kingdoms (Westergaard 1958; Ohno 1967; Bull 1983; Charlesworth 1991). Despite clear biological differences between these kingdoms, X and Y chromosomes in both lineages have undergone similar genetic changes. For example, in both groups the loss of recombination between X and Y chromosomes is associated with genetic degeneration and loss of genes on the Y chromosome (Hough *et al.* 2014; Bergero *et al.* 2015). In some species this has also led to the evoltion dosage compensation (Charlesworth 1996b; Muyle *et al.* 2012; Papadopulos *et al.* 2015), and changes to levels of X and Y genetic diversity (Ellegren 2011; Bachtrog 2013). The independent evolution of these phenomena in a broad range of species suggest that general evolutionary mechanisms may be involved, but inferring the causes of molecular evolution and patterns of polymorphism in regions that lack recombination is a difficult challenge for both theoreticians and experimentalists (Charlesworth 1978, 1996b; Charlesworth and Charlesworth 2000).

One fundamental difference between the X and Y chromosomes is that there are 1/3 as many Y-linked gene copies as X-linked ones in a diploid population. As a consequence, the neutrally-expectied effective population size (N_e) of the Y chromosome is predicted to be 1/4 that of the autosomes, whereas the N_e for the X chromosome should be 3/4 that of autosomes (assuming an equal number of reproducing females and males). Such differences in N_e are predicted to directly affect relative levels of neutral polymorphism maintained on these chromosomes, which is proportional to the product of N_e and the

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Monday 31st October, 2016%
[1]jhough@ucdavis.edu

neutral mutation rate, μ (Kimura 1984). Such variation in neutral polymorphism within the genome can have important consequences on patterns of DNA sequence evolution, including the effectiveness of both positive and negative selection (Charlesworth *et al.* 1987).

Several demographic factors are expected to modulate or accentuate differences in neutral polymorphism maintained on sex chromosomes (Ellegren 2009). These include population subdivision and sex-biased dispersal, deviations from a 1:1 breeding sex ratio, and high sex-specific variance in reproductive success (Caballero 1995; Charlesworth 2001; Laporte and Charlesworth 2002; Pool and Nielsen 2007). In theory, these processes all result in asymmetries in female vs male N_e, that could lead to different levels of neutral polymorphism on the sex chromosomes than are expected based on the autosomes ((Pool and Nielsen 2007)). As evidence for this, coalescent simulations have shown that models of male-biased migration during human out-of-Africa dispersal result in levels of X/A neutral diversity similar to what is observed (Keinan *et al.* 2009) (though it is worth noting that there is considerable variation in current estimates of X/A in humans, and other demographic scenarios have been suggested (Hammer *et al.* 2010; Bustamante and Ramachandran 2009)). In addition, disproportionate losses of sex-linked compared to autosomal variation arising from population bottleneck events have been inferred in chimpanzees and orangutans (Kaessmann *et al.* 2001; Fischer *et al.* 2006), the house mouse *Mus musculus* (Baines and Harr 2007), and Drosophila (Andolfatto 2001), highlighting the importance of demography in affecting patterns of sex-linked and autosomal neutral variation.

In addition to demographic factors, evolutionary theory predicts that neutral variability on Y chromosomes with suppressed recombination can be significantly reduced by Hill-Robertson Interference (HRI), which causes a reduction in N_e of neutral sites that are genetically linked to sites experiencing selection. The major types of HRI include (i) selection on strongly beneficial mutations ("selective sweeps") (Maynard Smith and Haigh 1974; Aquadro *et al.* 1994) (ii) selection against strongly deleterious mutations) (Charlesworth 1996a, 1994) (Charlesworth 1996b; Charlesworth and Charlesworth 2000), and (iii) intereference between weakly selected sites (weak selection Hill-Robertson interference). In all of these processes, linkage disequilibrium between selected and neutral sites results in a level of neutral variation below the level predicted in the absence of selection.

Consistent with an important role for Hill-Robertson interference in driving down diversity on the Y chromosome, studies examining Y chromosome variability in mammals (Hellborg and Ellegren 2004; Bachtrog 2013; Wilson Sayres *et al.* 2014) and Drosophila (McAllister and Charlesworth 1999; Bachtrog and Charlesworth 2000) have revealed that Y-linked diversity is considerably lower than predicted under models of neutral evolution, with evidence for linked selection playing a key role.

Despite widespread interest in determining the effects of recombination on patterns of genetic diversity (Ellegren 2011; Bachtrog 2013), we still know very little about the influence of linked selection on patterns nucleotide polymorphism in younger sex chromosome systems, where *de novo* recombination suppression evolved relatively recently (Charlesworth 2016). Background selection and selective sweeps are predicted to have the strongest effects during the earliest stages of Y-chromosome evolution, before Y-chromosomes have lost the majority of their genes. On the other hand, if gene silencing and Y inactivation occur very early during the process of Y chromosome degeneration, it is possible that the strength of linked selection is minimized rapidly even on relatively young Y chromosomes, and that much degeneration may occur neutrally following the silencing of Y-linked genes (cite backtrog NRG review).

To investigate the factors affecting nucleotide diversity in the early stages of sex chromosome evolution, we analyzed single nucleotide polymorphisms (SNPs) on sex chromosomes and autosomes in the plant *R. hastatulus* (Polygonaceae). This species is a dioecious annual with highly heteromorphic X and Y chromosomes that are estimated to have evolved approximately 15 MYA (Quesada del Bosque *et al.* 2011; Grabowska-Joachimiak *et al.* 2015; Navajas-Pérez *et al.* 2005), making sex chromosomes in this species over 100 million years younger than the well-studied mammalian sex chromosomes (Lahn and Page 1999; Ross *et al.* 2005). *Rumex hastatulus* has also received particular attention because of the unique occurrence of an intraspecific sex chromosome system, in which both XY and XY_1XY_2 males occur in geographically distinct 'chromosomal races' (Smith 1963). The XY_1XY_2 sex chromosome system in this species (the North Carolina race) is thought to have originated through an X-autosome fusion, with the XY system (the Texas race) maintaining the ancestral chromosome complement (Smith 1964). Despite the recent origin of sex chromosomes in both races, a recent study revealed that both the ancestral and neo-Y chromosomes have undergone significant gene loss and functional deterioration (Hough *et al.* 2014).

Of particular relevance to the present study, *R. hastatulus* populations have also been found to consistently exhibit female-biased sex ratios, with a mean sex ratio of $N_f/(N_m + N_f) = 0.6$ (Pickup and Barrett 2013). This is important because female-biased sex ratios are predicted to affect chromosomal N_e and therefore the level of neutral polymorphism maintained on sex chromosomes and autosomes (Ellegren 2009). Here, we test whether patterns of nucleotide diversity on sex chromosomes and autosomes in *R. hastatulus* are jointly consistent with neutral evolutionary processes - including the effects of sex ratio bias and variance in reproductive success - or whether the removal of linked neutral variation caused by positive and purifying selection has played a more significant role.

Materials and Methods

Population Samples and Sex-Linked Genes

We analyzed sex-linked and autosomal genes identified from Illumina RNA sequence data from 24 individuals (12 males and 12 females, with 1 male and 1 female from each of 12 populations; 6 populations of each sex chromosome race). Samples were collected in 2010 from throughout the native range of *R. hastatulus* (locations in Table S1), and plants were grown in the glasshouse from seeds collected from open-pollinated females. We extracted RNA from leaf tissue using Spectrum Plant Total RNA kits (Sigma-Aldrich). The isolation of mRNA and cDNA synthesis was conducted according to standard Illumina RNAseq procedures, with sequencing conducted on two Illumina HiSeq lanes with 150-bp end reads at the Genome Quebec Innovation Center. Reads from these 24 samples were mapped to the *R. hastatulus* reference transcriptome (Hough *et al.* 2014), and raw sequences are available from the GenBank Short Read Archive under accession no. SRP041588. Reads were mapped using the BurrowsWheeler Aligner (Li and

Durbin 2010), followed by Stampy (Lunter and Goodson 2011). We used Picard tools (http://picard.sourceforge.net) to modify mapping output into the format required for the Genome Analysis Toolkit (McKenna *et al.* 2010) variant calling software, and subsequently removed genes with low coverage (<10x) and Phred Quality Scores <20. The population samples analyzed here were previously reported in Hough et al. (2014), where they were used to validate the ascertainment of sex-linked genes identified through segregation analysis. Here, to consider sex linked genes that were identified in each of our sequenced population samples, we focused on the set of 460 X/Y genes for which the Y homolog was found in both the Texas and North Carolina races (i.e., X/Y genes where the Y copy was inferred to be on the Y_1 chromosome).

Autosomal Genes

In evaluating evidence for nucleotide diversity differences between X and Y chromosomes, it is important to distinguish between reduced Y-linked diversity, and the possibility that X-linked diversity is elevated above the level predicted from a neutral model. To do this, we normalized sex-linked diversity estimates by autosomal diversity, and compared empirical X/A and Y/A nucleotide diversity ratios to those predicted from neutral models and from simulations (described below). Because the criteria for identifying autosomal loci in Hough et al. (2014) were based on the occurrence of four segregating SNPs per locus, this set of genes is probably higher in diversity than the average autosome. Therefore, for the present analysis we incorporated the broader set of all non-sex linked (putatively autosomal) genes from our transcriptome data. We also filtered this set to remove genes that may have been sex-linked but were not identified as such by Hough et al.'s conservative ascertainment criteria. In particular, we removed: (i) any genes in which there was evidence for at least one SNP with a sex-linked segregation pattern, (ii) any genes with SNPs showing fixed heterozygosity in males and fixed homozygosity in females, (iii) any genes with less than 10X coverage or greater than 100X coverage from independently obtained genomic coverage data (to filter out duplicate genes or those with highly repetitive sequences), and (iv) any genes containing SNPs with large (>0.4) allele frequency differences between males and females. Finally, we removed genes with fewer than 100 synonymous sites following this filtering to avoid biasing our results toward genes that may have been particularly short due to assembly problems. This filtering resulted in a set of 12,356 and 11,350 autosomal genes in the Texas and North Carolina races, respectively.

Phasing X and Y alleles

To estimate polymorphism for the X and Y sequences separately, it is necessary to infer the phase of SNPs in sex-linked transcripts in males. In previous work, phasing alleles on *R. hastatulus* sex chromosomes was achieved using segregation analysis from a genetic cross. Here, to phase SNPs from population samples where such segregation data was unavailable, we used HAPCUT (Bansal and Bafna 2008), a maximum-cut based algorithm that reconstructs haplotypes using sequenced fragments (Illumina read data) from the two homologous chromosomes to output a list of phased haplotype blocks containing the SNP variants on each chromosome. Because the resulting haplotype blocks produced by HAPCUT contained SNPs that were phased relative to each other, but not designated to either the X or Y chromosome, we assigned individual variants to X or Y by independently identifying fixed X-Y differences with each haplotype block (i.e., sites in which all females were homozygous, and all males were heterozygous). Identifying such fixed differences within phased haplotype blocks enabled us to then infer the correct phase (X or Y) of the polymorphisms from HAPCUTs output. This was done by matching the phase of fixed X-Y differences with their neighboring polymorphic sites: when a fixed X-Y difference occurred in the same phased haplotype block as a polymorphic site, the polymorphic variants in that block were assigned to either X or Y based on the known phase of the fixed difference with which they were matched. SNPs that were identified outside of phased blocks, or in blocks without fixed X-Y differences, were recorded as missing data. Finally, we filtered out SNPs with coverage > 60, QUAL score > 60, and those within a distance of 10bp or less from indels. This procedure was conducted using a combination of Perl and Bash scripts, and resulted in fasta-formatted alignments of X and Y sequences for 372 sex-linked genes from the 24 individuals in our study.

We further validated the results of HAPCUTs allele phasing by comparing the accuracy of this method with the phasing-by-segregation method that was conducted in Hough et al. (2014). To do this, we first phased the sequence data from parents and their progeny using HAPCUTs algorithm (using the same parameters as for the population data), and then identified cases where SNPs were inferred on the Y chromosome by HAPCUT, but where the true level polymorphism was known to be zero. We identified 7 percent of sex-linked genes with phasing errors of this kind, or which were otherwise determined to have genotyping errors resulting in false SNP calls. This corresponds to a SNP error rate estimate of 1.7 x 10-4. Note that this rate is very low relative to population-based estimates of polymorphism on the X and autosomes (Table 1), and therefore should have minimal effects on our estimation of X/A polymorphism. However, because the rate is high relative to the expected level of true polymorphism on the Y-chromosome, we further filtered genes in which we found evidence for false-positive Y-polymorphisms arising from: (i) phasing errors caused by gene duplicates (more than two haplotypes), (ii) polymorphisms around indels, and (iii) genotyping errors caused by low Y-expression. This filtering was done by manually inspecting sequences in IGV (Robinson *et al.* 2011) and identifying each individual putative polymorphism on the Y chromosome.

Estimating nucleotide diversity on sex chromosomes and autosomes

For each locus in our analysis, we calculated Wattersons (1975) estimator of the population parameter $\theta = 4N_e\mu$, where N_e is the effective population size, and μ is the mutation rate (Watterson 1975), using a modified version of the Perl program Polymorphurama (Bachtrog and Andolfatto 2006). To compare sex-linked and autosomal loci, we calculated the average value of θ, weighted by the number of synonymous sites in each gene (**X** Figure 2; Table 1). We obtained 95 percent confidence intervals for our estimates of the X/A and Y/A diversity ratios by bootstrapping by gene using the BCa method (Efron and Tibshirani 1994) implemented in the Boot package in R (Canty and Ripley 2012), and calculating X/A and Y/A diversity on each iteration for 20000 replicates each. Bootstrapping was conducted on the final filtered set of 173 sex-linked, and 12355 autosomal genes from the Texas race, and separately for the 176 sex-linked and 11349 autosomal genes from the North Carolina

race. Note that the lack of recombination on the Y chromosome implies that assumptions about independence across loci are likely violated, suggesting that the true uncertainty in our Y/A diversity estimate may be wider than implied by bootstrapping across genes. To address this issue, we also used a maximum likelihood approach, implemented in a modified version of the MLHKA software (Wright and Charlesworth 2004) to independently estimate a credibility interval for the Y/A ratio using coalescent theory (Figure S1). Because of the thousands of genes involved, a likelihood method incorporating divergence to control for heterogeneity in mutation rate is not feasible, since this would involve maximizing the likelihood estimate of the mutation rate for each locus independently. Therefore, we applied a method that assumed no heterogeneity in mutation rate, no recombination between Y-linked genes, and free recombination between autosomal loci. Our model therefore had two parameters; theta for autosomal genes and the Y/A ratio. We varied both parameters across a grid of points and evaluated the likelihood, where the Y/A ratio varied from 0.001 to 1, and Watterson's theta on the autosomes varied from 0.001 to 0.01. We also tested whether estimates of diversity for the X chromosome calculated from phased sequences from females were consistent with estimates from phased sequences from males. As no significant difference was observed (**X** Figure? Table?), we report only results from females.

Finally, we tested for population substructure within and between the two sex chromosome races to control for the possibility that hidden substructure in our sampled populations could affect estimates of diversity. To do this, we constructed neighbor-joining trees of sex-linked sequences from all populations in our study. The analysis was conducted on an alignment of the X- and Y-linked genes from *R. hastatulus*, with orthologous autosomal sequences from the non-dioecious but closely related outgroup species *R. bucephalophorus* used to root the tree (Figure S2). We used the Neighbor-Joining method (Saitou and Nei 1987), with evolutionary distances computed using Maximum Composite Likelihood (Tamura *et al.* 2011). The inferred trees revealed strong support for Y-linked genes from the XY_1XY_2 ("North Carolina") race being paraphyletic (98% bootstrap support), with samples from two populations (hereafter the "SC sub-clade") forming a monophyletic group, and samples from Florida and Georgia (hereafter the "FL sub-clade") also forming a monophyletic group, that was more closely related to the XY ("Texas") race (Figure S2). In light of this evidence for significant population substructure, and to consider between-population differences, we estimated X and Y diversity in each of the three sub-clades (SC, FL, TX) separately. For model-based analyses, we focused on the population sample from the Texas XY race, given that this reflects the presumed ancestral karyotype and has the largest sample size.

Neutral predictions and the effect of sex ratio bias on diversity

To test whether the levels of diversity we observed on X, Y and autosomal chromosomes could be explained by the occurrence of biased reproductive sex ratios or high variance in reproductive success, we compared our empirical diversity estimates to predictions from a neutral model that jointly considers these two effects.

In particular, the effective population size for genes on each chromosome can be given by (Wright 1931):

$$N_{eA} = \frac{4N_mN_f}{N_m + N_f} \quad (1)$$

$$N_{eX} = \frac{9N_mN_f}{4N_m + 2N_f} \quad (2)$$

$$N_{eY} = \frac{N_m}{2} \quad (3)$$

Using these predictions, we calculated the expected X/A and Y/A ratios of diversity for the estimated *R. hastatulus* sex ratio $(\mathrm{N}_f/\mathrm{N}_m + N_f) = 0.6$ (Pickup and Barrett 2013), assuming that the level of neutral polymorphism is given by the product of the mutation rate and the effective population size, $\theta = 4N_e\mu$ (Watterson 1975). Sex-specific variance in reproductive success can lead to departures from this prediction, and so we also evaluated whether any ratio of male- to female- effective population sizes could explain our observed combination of X/A and Y/A diversity.

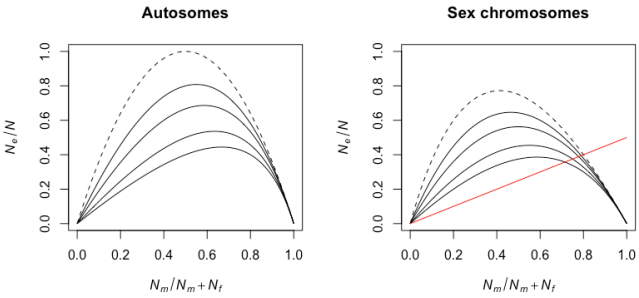

Figure 1 The relation between relative effective population size and sex ratio bias for genes on autosomes (**A**) and sex chromosomes (**B**) (Y chromosome in red). The sex ratio is shown as the proportion of males, $N_m/(N_f + N_m)$, where N_m and N_f are the effective number of breeding males and females, respectively, plotted against N_e/N, where $N = N_m + N_f$ and the N_e for sex chromosomes and autosomes is given in equations 1-3. Dotted curves show predictions in the standard neutral model, where both males and females produce Poisson-distributed offspring numbers and the chromosomal N_e's are given by equations 4-6 (Wright 1931). Solid curves correspond to increasing levels of variance in male reproductive success (Nomura 2002) (see Methods). Assuming $\theta = 4N_e\mu$ and equal neutral mutation rates among genes, the predicted N_e's are used to generate null predictions for X/A and Y/A ratios of diversity.

Simulations of purifying selection

To test whether our observed level of Y-chromosome diversity could be explained by the effects of linked purifying selection, we followed the approach used in (Wilson Sayres *et al.* 2014) and conducted forward-time simulations of a nonrecombining Y chromosome using the software SFSCODE (Hernandez 2008) and compared our empirical diversity estimates on the Y with those from simulations. This approach was preferred over using analytical predictions of background selection primarily because the equilibrium background selection model over-predicts the reduction in diversity when there are many linked sites under selection (Kaiser and Charlesworth 2009), as is expected to be the case for large Y chromosomes that lack crossing over.

Table 1 Estimates of neutral diversity by race on *R. hastatulus* sex chromosomes and autosomes.

	Texas		South Carolina		Florida	
chromosome	θ	θ/θ_A	θ	θ/θ_A	θ	θ/θ_A
A	0.006	1	0.006	1	0.005	1
X	0.0047	0.85	0.002	0.33	0.0047	0.37
Y	10^{-4}	0.002	10^{-4}	0.002	10^{-4}	0.002

To parameterize the forward simulations, we estimated the distribution of fitness effects of deleterious mutations (DFE) using our polymorphism data on the X-chromosome with the program DFE-alpha (Keightley and Eyre-Walker ref). This program fits a gamma distribution of selection coefficients to the observed frequency distribution of nonsynonymous and synonymous polymorphism from the population genetic analyses above. We used this estimated gamma distribution in our forward simulations. To examine the expected reduction in diversity, we varied the number of sites under selection, and obtained an approximate likelihood surface for our observed data based on the proportion of simulations that matched the observed estimate of $\pi_{synonymous}$. To make the simulation output comparable to our data, we initialized the simulations with our empirically estimated autosomal θ, adjusted for the expected neutral reduction in effective population size on the Y chromosome under a sex ratio of Nf/(Nf+Nm)=0.6, and we sampled 6 haploid chromosomes per simulation. Simulated sequences contained 45,331 of linked neutral sequence from which we calculated diversity, matching the number of synonymous sites sampled in our empirical analyses on the Y chromosome (see Supporting Information for simulation commands).

Results and Discussion

Y-chromosome diversity in R. hastatulus is very low

Our analysis revealed that, for all sex chromosome races, diversity on the *R. hastatulus* Y chromosome is significantly lower than expected under neutrality, with estimates indicating Y/A=0.02, which is 12.5 fold lower than the standard neutral prediction of Y/A = 0.25 (P<0.0001). In the XY chromosome race, we also observe that the Y chromosome shows a 40-fold lower than mean diversity on the X chromosome (Table 1). Note that by normalizing X and Y diversity by autosomal diversity, our results indicate that the X-Y difference we observed was not due to an elevation of X chromosome diversity, but rather a Y-specific reduction. Conceivably, such low diversity on the Y could arise from a low mutation rate on the Y chromosome, or a lower mutation rate in males compared to females. However, these possibilities are unlikely because there is no evidence that the number of synonymous mutations in X and Y lineages, estimated by both parsimony and maximum likelihood, are significantly different (Hough *et al.* 2014). Similarly, our estimates of weighted average synonymous substitution rate between R. hastatulus and an outgroup species R. bucephalophorus from Hough et al(ref) reveal comparable levels of synonymous divergence at sex-linked (0.2016) and autosomal genes (0.219).

Although our sampling of each *R. hastatulus* sub-clade is limited, the discovery of three phylogenetically distinct monophyletic groups is interesting because it suggests the possibility that introgression occurred between the ancestral Texas (XY race) and the derived North Carolina (XY_1Y_2) race, leading to a derived XY_1Y_2 sub-clade. As the North Carolina and Texas races are known to be inter-fertile (Smith 1964), we suggest that the SC sub-clade inferred here likely originated through hybridization between a female from the FL clade harboring the X-autosome fusion, and a male from the XY Texas race. Notably, when we pooled the samples from each sub-clade together, the estimated level of Y diversity was significantly higher in the pooled data, highlighting the presence of strong substructure between these groups (Figure S2).

Our results also indicate a significant reduction in X/A diversity in the derived SC and FL sub-clades of the North Carolina race ($X/A_{FL} = 0.33$ and $X/A_{SC} = 0.37$) compared to the Texas race ($X/A_{TX} = 0.85$) (Figure 2). Although not expected, this reduction in diversity may be associated with the recent origin of the XY_1Y_2 sex chromosome system, which is thought to have originated through an X-autosome fusion involving the ancestral 3rd chromosome in the Texas race (Smith 1964). Evidence supporting this autosomal origin was recently obtained by (Grabowska-Joachimiak *et al.* 2015), who reported that the ancestral third chromosome in the Texas race carries the 5S rDNA locus, which is now found on both the neo-X and the Y_2 sex chromosomes in the derived North Carolina race. If recent positive selection was involved in driving the evolution of this X-A fusion, which theory suggests can drive the evolution of such fusions (Charlesworth and Charlesworth 1980), then the formerly autosomal segment on the X chromosome in the XY_1Y_2 sub-clades may have experienced a strong selective sweep, resulting in reduced X-linked diversity in the derived XY_1Y_2 sub-clades. However, this would require extensive recombination suppression between the fused and unfused X chromosomes, and it is also possible that sex-specific demographic history has driven these patterns. It will be important for future work to investigate in more detail the factors driving the establishment of the X-autosome fusion in this species, and how they might impact patterns of X-linked neutral diversity.

In contrast to the X chromosome, however, our data indicate a strong and consistent diversity reduction on the Y chromosome: an approximately 50 fold reduction compared to the mean θ_{Aut}. We next consider several possible models - neutral and selective - that might explain this reduction.

Female biased sex ratio and high variance in male fitness

The occurrence of female-biased sex ratios in this species has been predicted to lower Y diversity due to its effects on reducing male N_e and therefore the neutrally expected N_e of the Y chromosome (**X** ref). This reduction N_e on the Y is expected to be further accented if there is high variance in male reproductive success (Figure 1), which is not unusual in annual plants such as *R. hastatulus* that commonly exhibit extensive phenotypic plasticity in plant size and flower production (**X** Harper 1977). Moreover, given that male plants in this wind-pollinated species produce large amounts of pollen, and that female flowers are uniovulate, we expected that there is strong competition among males to fertilize females.

In common with most flowering plants we do not have marker-based estimates of the variance in male reproductive success in *R. hastatulus*. However, by comparing our empirical estimates of diversity to predictions from models that jointly predict the effects on diversity of sex ratio bias and male repro-

ductive variance, we evaluated whether these effects could explain the level of Y/A diversity that we observed (see Methods). Conditioning on estimates of sex ratio bias in *R. hastatulus* that have been estimated, ranging from $N_m/(N_m+N_f) = 0.4$ to $N_m/(N_m+N_f) = 0.35$ (Pickup and Barrett 2013), the predicted Y/A diversity ratio is approximately 0.2 (Table1, Figure 3). This is significantly lower than our estimated mean Y/A ratio of 0.02 ($P < 0.0001$), and remains significant even if we consider an upper bound estimate obtained from our likelihood based estimated of the confidence interval. This suggests that the sex ratio effect alone is insufficient to explain our data under the standard sex ratio model. (**X** I think some of this can be written better...)

Assuming that there is extensive variance in male reproductive success (**X**), the predicted ratios of Y/A diversity (**X** supp) were also significantly higher than our estimates for the empirically estimated sex ratio (**X** table) . We also find, however, that purely neutral models in which the sex ratio was highly female-biased (**X**), with level of variance in male reproductive success on the order of (**X**), predicted a Y/A ratio that could not be rejected (**X**). However, because a highly females biased sex ratio is expected to increase the X/A ratio as well, these models simultaneously predicted a range of X/A ratios that were significantly different from what we observed (**X** Figure). Thus, our results indicate that the combined effects of sex ratio bias and variance in reproductive success cannot jointly explain our observed levels of X, Y, and autosomal diversity.

Background Selection and Selective Sweeps

It is worth noting that our results only apply to the Y_1 chromosome, as estimates of diversity were calculated for sex-linked genes that were shared between the XX/XY and the XX/XY_1Y_2 sex chromosome systems. Previous work found that both the Y_1 and the more recently evolved "neo" Y_2 chromosomes exhibited signs of genetic degeneration, including gene loss, loss of expression, and an accumulation of amino acid-changing mutations (Hough *et al.* 2014), but we cannot say from the present study whether the neo-Y chromosome has also undergone a reduction in diversity. However, the extensive reduction in diversity estimated on the Y_1 chromosome occurs in each of the three *R. hastatulus* sub-clades (**X** Figure 2; Table 1; Figure S2), suggesting that this effect is not population-specific.

Conclusions

The observation of widespread degeneration and diversity loss on Y chromosomes illustrates the importance of recombination for maintaining fitness and the genetic variability needed for adaptation (Maynard Smith 1978; Kondrashov 19 93; Barton Charlesworth 19 9 8). The accumulation of deleterious mutations and extensive loss of diversity on a young plant Y chromosome revealed by our study provides a clear example of the reduced ability of selection to eliminate deleterious alleles from large genomic regions that lack recombination.

Acknowledgments

Literature Cited

Andolfatto, P., 2001 Contrasting patterns of x-linked and autosomal nucleotide variation in drosophila melanogaster and drosophila simulans. Molecular Biology and Evolution **18**: 279–290.

Aquadro, C. F., D. J. Begun, and E. C. Kindahl, 1994 Selection, recombination, and dna polymorphism in drosophila. In *Non-neutral evolution*, pp. 46–56, Springer.

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.

Bachtrog, D. and B. Charlesworth, 2000 Reduced levels of microsatellite variability on the neo-y chromosome of drosophila miranda. Current Biology **10**: 1025–1031.

Baines, J. F. and B. Harr, 2007 Reduced x-linked diversity in derived populations of house mice. Genetics **175**: 1911–1921.

Bansal, V. and V. Bafna, 2008 Hapcut: an efficient and accurate algorithm for the haplotype assembly problem. Bioinformatics **24**: i153–i159.

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**.

Charlesworth, B., 1978 Model for evolution of y chromosomes 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., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genetical research **63**: 213–227.

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 chromosomes. Philosophical Transactions of the Royal Society of London B: Biological Sciences **355**: 1563–1572.

Charlesworth, B., J. Coyne, and N. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. American Naturalist pp. 113–146.

Charlesworth, D., 2016 Plant sex chromosomes. Annual review of plant biology **67**: 397–420.

Charlesworth, D. and B. Charlesworth, 1980 Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genetical research **35**: 205–214.

Efron, B. and R. J. Tibshirani, 1994 *An introduction to the bootstrap*. CRC press.

Ellegren, H., 2009 The different levels of genetic diversity in

sex chromosomes and autosomes. Trends in Genetics **25**: 278–284.

Ellegren, H., 2011 Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nature Reviews Genetics **12**: 157–166.

Fischer, A., J. Pollack, O. Thalmann, B. Nickel, and S. Pääbo, 2006 Demographic history and genetic differentiation in apes. Current Biology **16**: 1133–1138.

Grabowska-Joachimiak, A., A. Kula, T. Książczyk, J. Chojnicka, E. Sliwinska, and A. J. Joachimiak, 2015 Chromosome landmarks and autosome-sex chromosome translocations in rumex hastatulus, a plant with xx/xy1y2 sex chromosome system. Chromosome Research **23**: 187–197.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, M. P. Cox, and J. D. Wall, 2010 The ratio of human x chromosome to autosome diversity is positively correlated with genetic distance from genes. Nature genetics **42**: 830–831.

Hellborg, L. and H. Ellegren, 2004 Low levels of nucleotide diversity in mammalian y chromosomes. Molecular Biology and Evolution **21**: 158–163.

Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. Bioinformatics **24**: 2786–2787.

Hough, J., J. D. Hollister, W. Wang, S. C. Barrett, and S. I. Wright, 2014 Genetic degeneration of old and young y chromosomes in the flowering plant rumex hastatulus. Proceedings of the National Academy of Sciences **111**: 7713–7718.

Kaessmann, H., V. Wiebe, G. Weiss, and S. Pääbo, 2001 Great ape dna sequences reveal a reduced diversity and an expansion in humans. Nature genetics **27**: 155–156.

Kaiser, V. B. and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends in Genetics **25**: 9–12.

Keinan, A., J. C. Mullikin, N. Patterson, and D. Reich, 2009 Accelerated genetic drift on chromosome x during the human dispersal out of africa. Nature genetics **41**: 66–70.

Kimura, M., 1984 *The neutral theory of molecular evolution*. Cambridge University Press.

Lahn, B. T. and D. C. Page, 1999 Four evolutionary strata on the human x chromosome. Science **286**: 964–967.

Laporte, V. and B. Charlesworth, 2002 Effective population size and population subdivision in demographically structured populations. Genetics **162**: 501–519.

Li, H. and R. Durbin, 2010 Fast and accurate long-read alignment with burrows–wheeler transform. Bioinformatics **26**: 589–595.

Lunter, G. and M. Goodson, 2011 Stampy: a statistical algorithm for sensitive and fast mapping of illumina sequence reads. Genome research **21**: 936–939.

Maynard Smith, J. and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genetical research **23**: 23–35.

McAllister, B. F. and B. Charlesworth, 1999 Reduced sequence variability on the neoy chromosome of drosophila americana americana. Genetics **153**: 221–233.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, *et al.*, 2010 The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. Genome research **20**: 1297–1303.

Muyle, A., N. Zemp, C. Deschamps, S. Mousset, A. Widmer, and G. A. Marais, 2012 Rapid de novo evolution of x chromosome dosage compensation in silene latifolia, a plant with young sex chromosomes. PLoS Biol **10**: e1001308.

Navajas-Pérez, R., R. de la Herrán, G. L. González, M. Jamilena, R. Lozano, C. R. Rejón, M. R. Rejón, and M. A. Garrido-Ramos, 2005 The evolution of reproductive systems and sex-determining mechanisms within rumex (polygonaceae) inferred from nuclear and chloroplastidial sequence data. Molecular Biology and Evolution **22**: 1929–1939.

Nomura, T., 2002 Effective size of populations with unequal sex ratio and variation in mating success. Journal of Animal Breeding and Genetics **119**: 297–310.

Ohno, S., 1967 *Sex chromosomes and sex-linked genes*, volume 1. Springer-Verlag.

Papadopulos, A. S., M. Chester, K. Ridout, and D. A. Filatov, 2015 Rapid y degeneration and dosage compensation in plant sex chromosomes. Proceedings of the National Academy of Sciences **112**: 13021–13026.

Pickup, M. and S. C. Barrett, 2013 The influence of demography and local mating environment on sex ratios in a wind-pollinated dioecious plant. Ecology and evolution **3**: 629–639.

Pool, J. E. and R. Nielsen, 2007 Population size changes reshape genomic patterns of diversity. Evolution **61**: 3001–3006.

Quesada del Bosque, M., R. Navajas-Pérez, J. Panero, A. Fernández-González, and M. Garrido-Ramos, 2011 A satellite dna evolutionary analysis in the north american endemic dioecious plant rumex hastatulus (polygonaceae). Genome **54**: 253–260.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov, 2011 Integrative genomics viewer. Nature biotechnology **29**: 24–26.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, D. Muzny, M. Platzer, G. R. Howell, C. Burrows, C. P. Bird, *et al.*, 2005 The dna sequence of the human x chromosome. Nature **434**: 325–337.

Saitou, N. and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution **4**: 406–425.

Smith, B. W., 1963 The mechanism of sex determination in rumex hastatulus. Genetics **48**: 1265.

Smith, B. W., 1964 The evolving karyotype of rumex hastatulus. Evolution pp. 93–104.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, 2011 Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution **28**: 2731–2739.

Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. Theoretical population biology **7**: 256–276.

Westergaard, M., 1958 The mechanism of sex determination in dioecious flowering plants. Advances in genetics **9**: 217–281.

Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen, 2014 Natural selection reduced diversity on human y chromosomes. PLoS Genet **10**: e1004064.

Wright, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.

Wright, S. I. and B. Charlesworth, 2004 The hka test revisited. Genetics **168**: 1071–1076.