Reduced Nucleotide Diversity 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, both positive and negative selection can remove genetically linked neutral variation, resulting in a reduction in N_e and the level of polymorphism on Y compared to X chromosomes or autosomes. However, non-selective factors including biased sex ratios and high variance in male fitness can also reduce Y-linked N_e, making it difficult infer the causes of low Y-chromosome diversity. Here, we investigate the factors affecting X- and Y-linked polymorphism during the early stages of plant sex chromosome evolution in *Rumex hastatulus* (Polygonaceae). Strikingly, we find that nucleotide diversity on the Y chromosome is 40 fold lower than on the X chromosome. We demonstrate that the magnitude of this reduction is not consistent with a purely neutral model based on the observed occurrence of female biased sex ratios in this species, or by reduced male N_e caused by high variance in male fitness. Rather, using forward simulations, we show that the diversity reduction on the Y is consistent with the predicted effects of purifying selection removing deleterious mutations and linked neutral variation. Given the recent origin of *R. hastatulus* sex chromosomes, our results indicate that the low level of sequence diversity commonly observed on ancient Y chromosomes might evolve soon after sex chromosomes originate, with background selection playing an important role during the early stages of their evolution.

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 numerous biological differences between the kingdoms, X and Y chromosomes in both lineages have undergone similar genetic changes, suggesting that general evolutionary mechanisms might drive sex chromosome divergence (Charlesworth 1978, 1996b; Charlesworth and Charlesworth 2000). Investigating the shared features of plant and animal sex chromosomes, including the loss of recombination (Bergero and Charlesworth 2009), the degeneration of

Y chromosomes (Hough *et al.* 2014; Bergero *et al.* 2015), and the evolution of dosage compensation (Muyle *et al.* 2012; Papadopulos *et al.* 2015), thus provides an informative context for investigating the evolutionary forces driving sex chromosome evolution.

One fundamental difference between sex chromosomes and autosomes is their respective copy number in a population; whereas autosomes in diploid species are present in two copies in each sex, the X chromosome is present in two copies in females and only one copy in males. As a consequence, the effective population size (N_e) of the X chromosome is predicted to be 3/4 that of autosomes, whereas the N_e of the male-limited Y chromosome is expected to equal 1/4 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 neutral mutation rate, μ

(Kimura 1984). This variation in neutral polymorphism is in turn expected to 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 can modulate or accentuate differences in polymorphism levels maintained on sex chromosomes (Ellegren 2009). These include population subdivision and sex-biased dispersal, deviations from a 1:1 breeding sex ratio, and high variance in reproductive success (Caballero 1995; Charlesworth 2001; Laporte and Charlesworth 2002; Pool and Nielsen 2007). For example, studies comparing levels of variability on the X chromosome and autosomes in humans have found evidence that sex-biased dispersal has reduced levels of X-to-autosome diversity below the neutrally expected level of 0.75 (Keinan *et al.* 2009), though there is considerable variation in estimates of the X/A ratio in humans, and other demographic processes have also been suggested (Hammer *et al.* 2010; Bustamante and Ramachandran 2009). Theoretical work also indicates that population bottlenecks can lead to disproportionate decreases in sex-linked compared to autosomal variation (Pool and Nielsen 2007), a prediction that has been supported in a diverse array of taxa, including chimpanzees and orangutans (Kaessmann *et al.* 2001; Fischer *et al.* 2006), the house mouse *Mus musculus* (Baines and Harr 2007), Drosophila (Andolfatto 2001), and humans (Keinan *et al.* 2009).

In addition to demographic factors, evolutionary theory predicts that genetic variability on sex chromosomes should be affected by the evolution of suppressed X-Y recombination, which lowers the coalescent N_e for Y-linked genes and increases their vulnerability to the effects of selective sweeps (Maynard Smith and Haigh 1974; Aquadro *et al.* 1994) and background selection (Charlesworth 1996a, 1994). These processes can be viewed as examples of the Hill–Robertson (HR) effect, in which the effective population size for a given chromosomal region depends strongly on the rate of recombination, such that selection at a focal site is not independent of selection at nearby sites (Hill and Robertson 1966). Moreover, both background selection and selective sweeps predict a reduction in N_e and, as a result, reduced levels of neutral polymorphism, selection efficacy, and rate of adaptation (Comeron *et al.* 2008). Although the HR effect can occur throughout the genome (McVean and Charlesworth 2000), large genomic regions in which recombination is suppressed, such as the Y chromosome, are expected to experience the most severe reductions in neutral diversity due to chromosome-wide linkage (Charlesworth 1996b; Charlesworth and Charlesworth 2000; Bachtrog 2013). Consistent with this, studies examining Y chromosome variability in mammals (Hellborg and Ellegren 2004; Wilson Sayres *et al.* 2014) and Drosophila (McAllister and Charlesworth 1999; Bachtrog and Charlesworth 2000) have found that Y-linked diversity is considerably lower than predicted under models of neutral evolution, suggesting that the HR effect plays a key role determining patterns of nucleotide diversity during sex chromosome evolution across a broad range of animal taxa.

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 the HR effect on patterns nucleotide polymorphism in younger sex chromosome systems, where *de novo* recombination suppression evolved relatively recently (Charlesworth 2016). Recent studies of young plant sex chromosome systems have revealed that several canonical features of X and Y chromosome evolution

are shared between plants and animals, including the genetic deterioration of the Y chromosome (Bergero *et al.* 2015; Hough *et al.* 2014), the occurrence of 'evolutionary strata' (Bergero and Charlesworth 2009), and the evolution of dosage compensation (Muyle *et al.* 2012; Papadopulos *et al.* 2015). Nonetheless, except for a small number of genes in *Silene latifolia* (Filatov *et al.* 2001; Qiu *et al.* 2010), the associated changes in genetic diversity that are predicted from evolutionary models of X- and Y-chromosome evolution have not been widely studied in plant sex chromosomes. Given that 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 (Bachtrog 2008), studying recently evolved sex chromosomes provides an important test of the temporal dynamics of the HR effect.

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 genetic degeneration(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 $N_m/N_f = 0.6$ (Pickup and Barrett 2013). This is important because female-biased sex ratios are predicted to affect chromosomal N_e and therefor the level of genetic diversity maintained on sex chromosomes and autosomes (Ellegren 2009). Here, we investigated whether the patterns of diversity we observed on sex chromosomes and autosomes were jointly consistent with predictions from models of sex ratio bias and variance in reproductive success, or whether linked selection (background selection and selective sweeps) has played a greater 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 Illu-

mina 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 Gen-Bank Short Read Archive under accession no. SRP041588. Reads were mapped using the Burrows–Wheeler 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, we focused on the set of 460 sex-linked genes for which the Y homolog was inferred to be on the y_1 chromosome (i.e., sex-linked genes that are shared between the Texas and North Carolina races).

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 a neutral model (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 (punitively 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 the conservative ascertainment criteria in Hough et al. (2014). 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 HAPCUT's output. This was done by matching the phase of fixed X-Y differences with their neighboring polymorphic sites - i.e., 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 HAPCUT's allele phasing by comparing the accuracy of this method with the phasing-by-segregation method that was conducted using data from a genetic cross in Hough et al. (2014). To do this, we first phased the sequence data from parents and their progeny using HAPCUT's algorithm (using the same parameters as for the population data). In particular, we looked for cases where SNPs were inferred on the Y chromosome by HAPCUT, but where the 'true' level polymorphism was inferred to be zero by phasing-by-segregation. We identified 7 percent of sex-linked genes with phasing errors of this kind, or that were determined to have genotyping errors resulting in a false SNP calls on the Y chromosome, with a SNP error rate estimate of 1.7 x 10-4. This rate is very low relative to the 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 this 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 polymorphism levels on sex chromosomes and autosomes

For each locus in our analysis, we calculated Watterson's (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 of independent loci are likely violated such that the true uncertainty in our Y/A diversity estimate obtained by this method is likely underestimated. To address this issue, we used a maximum likelihood approach, implemented a modified version of the MLHKA software (Wright and Charlesworth 2004), to independently estimate a credibility interval for the Y/A ratio (Figure S1). 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 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 conducted a phylogenetic analysis of sex-linked sequences from all populations in our study. The analysis was conducted on an alignment of 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). The phylogeny was inferred using the Neighbor-Joining method (Saitou and Nei 1987), with evolutionary distances computed using Maximum Composite Likelihood (Tamura *et al.* 2011). The inferred sequence phylogeny revealed strong support for Y-linked genes from the XY_1XY_2 ("North Carolina") race being paraphyletic (98 percent 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, but more closely related to the XY (Texas) race (Figure S2). In light of this evidence for population substructure, we used the inferred phylogeny as the basis for estimating X and Y diversity.

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 estimates of diversity to predictions from a neutral model that jointly considers these two effects. Here, following Nomura (2002), the effective population sizes for autosomal loci is given by:

$$N_{eA} = \frac{4N_mN_f}{N_m(1+C_f^2) + N_f(1+C_m^2)} \tag{1}$$

where N_m and N_f are effective population sizes for males and females, respectively, and C_i^2 is the coefficient of variation in mating success for the ith sex, defined as the ratio of the standard deviation to the mean number of mates σ_i/μ_{di}. Similarly, for X and Y linked loci,

$$N_{eX} = \frac{9N_mN_f}{4N_m(1+C_f^2) + 2N_f(1+C_m^2)} \tag{2}$$

and

$$N_{eY} = \frac{N_m(1+C_f^2)}{2} \tag{3}$$

With with no variation in mating success in either sex ($C_m^2 = C_f^2 = 0$), equations 1-3 reduce to the familiar expressions from (Wright 1931):

$$N_{eA} = \frac{4N_mN_f}{N_m + N_f} \tag{4}$$

$$N_{eX} = \frac{9N_mN_f}{4N_m + 2N_f} \tag{5}$$

$$N_{eY} = \frac{N_m}{2} \tag{6}$$

To consider the joint effects of biased sex ratios and variance in male reproductive success, our neutrally-expected levels of diversity were calculated using equations 1-3, assuming no variance in female mating (e.g., $C_f^2 = 0$), and we considered a range of values for σ_m: 0 (no variance), 1/2 (Poisson/2) 1 (Poisson distribution of offspring), 2 (Poisson*2), and 3 (Poisson*3). 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; **?**), Figure 1 shows the predicted effective population size of autosomes and sex chromosomes as a function of the sex ratio and variance in male reproductive success. Using these neutral predictions, we compared our empirically estimated diversity levels to normalized X/A and Y/A ratios of diversity. Note that we have only considered variance in male reproductive success here because this is most likely to occur in wind pollinated plant like *R. hastatulus*, where males produce copious amounts of pollen and compete to fertilize uniovulate females (**X** ref).
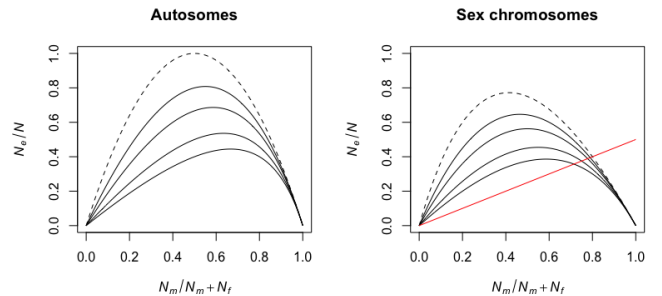

Figure 1 The relation between relative effective population size and sex ratio bias for genes on autosomes (**A**) and sex chromosomes (**B**) (Y chromosome in red). The sex ratio is shown as the proportion of males, $N_m/(N_f + N_m)$, where N_m and N_f are the effective number of breeding males and 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 positive and purifying selection

To test whether our observed level of Y-chromosome diversity could be explained by selection, we conducted two sets of forward-time simulations using the software SFSCODE (Hernandez 2008): (i) Purifying selection acting to remove deleterious mutations and linked neutral variation (i.e., background selection), and (ii) Positive selection acting on beneficial mutations and reducing variation through selective sweeps (refs **X**). For purifying selection simulations, we assumed that selection coefficients followed a gamma distribution, and we considered a range of values for the mean selection coefficient (*s*) by varying the scale parameter, *k* (range **X**) from (**X**) to **X** and keeping the shape parameter, **X**, constant (see Supporting Information for simulation code). To make the simulation output comparable to our data, we initialized simulations with our empirically estimated value of θ (estimated from autosomes), and sampled 8 chromosomes per simulation from 24 populations. For simulating positive selection, we used **X** [parameters here]. We ran 1000 replicates for each model and calculated the average θ_{sim} and the corresponding Y/A ratio of neutral diversity under the both sets of models separately (purifying and positive selection). We then tested the fit of our simulated diversity output to our empirical data (**X** new figure) by calculating the proportion of simulation replicates for which θ_{sim} was not significantly different from $\theta_{observed}$, using a two-sided P value. Commands for running the simulations are provided in the Supporting Information.

Results and Discussion

Y-chromosome diversity in R. hastatulus is very low

Our analysis revealed that 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 $Y/A = 0.25 (P < 0.0001)$, and a 40-fold reduction compared to 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 can be rejected 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). It is worth noting that our results only apply to the Y_1 chromosome, however, as estimates of diversity were calculated using the set of sex-linked genes that were shared between the XX/XY and the XX/XY_1Y_2 sex chromosome races. Previous work investigating patterns of genetic degeneration in both systems found that both the Y_1 and the more recently evolved "neo" Y_2 chromosomes have undergone classic signs of degeneration, including gene loss, loss of expression, and an accumulation of amino acid-changing mutations, but we cannot say at present whether the neo-Y has also undergone a significant reduction in diversity. Nonetheless, our results indicate that the pattern of reduced diversity on the Y_1 chromosome occurs in each of the *R. hastatulus* sub-clades that were inferred to be monophyletic (**X** Figure 2; Table 1; Figure S1,S2), suggesting that this effect is not population-specific.

Although our sampling of each *R. hastatulus* sub-clade is limited, the discovery of three phylogenetically distinct mono-

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

chromosome	Texas θ	Texas θ/θ_A	South Carolina θ	South Carolina θ/θ_A	Florida θ	Florida 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

phyletic groups is interesting as 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).

Female biased sex ratios and variance in male fitness

A highly reduced level of Y chromosome variability is a predicted outcome of models Y-chromosome evolution based on the removal of deleterious mutations and linked neutral variants (e.g., background selection), ...and several cases.... but the occurrence of female-biased sex ratios in this species has also been predicted to lower Y diversity. Moreover, a further reduction in Y diversity could occur because of high variance in male reproductive success (Figure 1). This is not unusual in annual species such as *R. hastatulus*s, which commonly exhibit extensive phenotypic plasticity in plant size and flower production (**X** Harper 1977). Conceivably, such plastic differences among plants, owing to spatial heterogeneity in resources, could influence the intensity of male-male competition, and thus affect variance in male reproductive success. Given that male plants in this wind-pollinated species produce large amounts of pollen, and that female flowers are uniovulate, one might expect strong competition among males.

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 modeling...we evaluated whether such an effect could result in a level of Y/A diversity similar to what we observed by comparing our results to predictions from a model for the joint effects of sex ratio bias and male reproductive variance (see Methods). Estimates of sex ratio bias in *R. hastatulus* have been estimated, and they are consistently female-biased, ranging from $N_m/(N_m + N_f) = 0.4$ to $N_m/(N_m + N_f) = 0.35$ (with a mean of 0.62; N=46 (Pickup and Barrett 2013).The Y/A diversity ratio predicted from this sex ratio, assuming a Poisson distribution of offspring numbers, is approximately 0.2 (Table1, Figure 3). Our estimated mean Y/A ratio, however, is 0.02, which is significantly lower ($P < 0.0001$). Indeed, even for higher values for the variance in male reproductive success (**X**), the corresponding predictions for Y/A are all significantly higher than our observed ratio (**X** table). Finally, neutral models that could not be rejected, i.e., in which a highly

female-biased sex ratio (**X**) and a high variance in male reproductive success predicted an Y/A ratio that fit our data (**X**), simultaneously predicts X/A ratios that are 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 and Y diversity in *R. hastatulus*.

Background Selection and Selective Sweeps

Conclusions

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., 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.

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. and D. Charlesworth, 2009 The evolution of restricted recombination in sex chromosomes. Trends in Ecology & Evolution **24**: 94–102.

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.

Comeron, J. M., A. Williford, and R. Kliman, 2008 The hill–robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity **100**: 19–31.

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.

Filatov, D. A., V. Laporte, C. Vitte, and D. Charlesworth, 2001 Dna diversity in sex-linked and autosomal genes of the plant species silene latifolia and silene dioica. Molecular Biology and Evolution **18**: 1442–1454.

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.

Hill, W. G. and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genetical research **8**: 269–294.

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.

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.

McVean, G. A. and B. Charlesworth, 2000 The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155**: 929–944.

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.

Qiu, S., R. Bergero, A. Forrest, V. B. Kaiser, and D. Charlesworth, 2010 Nucleotide diversity in silene latifolia autosomal and sex-linked genes. Proceedings of the Royal Society of London B: Biological Sciences p. rspb20100606.

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.